

Appendix A

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:35:18 ; Search time 18 Seconds  
(without alignments)  
608.735 Million cell updates/sec

Title: US-09-972-970-4  
Perfect score: 1283  
Sequence: 1 MFQKHQFQEPVGGCKYF.....RAPYTPKAWASLRSGCRTT 233

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	885	69.0	268	1 T4S9_HUMAN	O60628 homo sapien
2	324	25.3	294	1 TNE7_HUMAN	O95858 homo sapien
3	251	19.6	238	1 T4S7_MOUSE	O94833 mus musculus
4	243	18.9	238	1 T4S7_HUMAN	O14817 homo sapien
5	242	18.9	253	1 C151_RAT	O94246 rattus norv
6	240.5	18.7	245	1 T4S6_HUMAN	O43657 mus musculus
7	240	18.7	249	1 T4S2_HUMAN	P41732 homo sapien
8	239.5	18.7	245	1 T4S6_MOUSE	O70401 mus musculus
9	239	18.6	218	1 CD53_MOUSE	O61451 mus musculus
10	239	18.6	253	1 C151_CERAE	Q9MYM2 cercopithec
11	239	18.6	253	1 C151_MOUSE	O35566 mus musculus
12	236	18.4	239	1 TNE5_HUMAN	O75954 homo sapien
13	236	18.4	249	1 T4S2_MOUSE	O62283 mus musculus
14	235.5	18.4	219	1 CD53_HUMAN	P19397 homo sapien
15	234	18.2	253	1 C151_HUMAN	P49509 homo sapien
16	233.5	18.2	218	1 CD53_RAT	F24485 rattus norv
17	231	18.0	237	1 T4S3_HUMAN	P19075 homo sapien
18	228	17.8	267	1 CD82_HUMAN	P27701 homo sapien
19	220	17.1	237	1 CD63_MOUSE	P41731 mus musculus
20	217	16.9	266	1 CD63_MOUSE	P40237 mus musculus
21	215.5	16.8	236	1 CD63_BOVIN	O94832 mus musculus
22	215.5	16.8	266	1 CD82_RAT	O70352 rattus norv
23	214.5	16.7	225	1 CD9_PIG	O8WMQ3 sus scrofa
24	214	16.7	237	1 CD63_HUMAN	P08962 homo sapien
25	214	16.7	237	1 CD63_RABIT	Q28709 oryctolagus
26	211	16.4	237	1 CD63_RAT	P28648 rattus norv
27	210	16.4	236	1 CD81_MOUSE	P33762 mus musculus
28	208	16.2	225	1 CD9_MOUSE	P40240 mus musculus
29	208	16.2	236	1 CD81_RAT	Q62745 rattus norv
30	208	16.2	236	1 CD81_SAGOE	Q9N0J9 saguinus oe
31	207	16.1	236	1 CD81_CERAE	O97703 cercopithec
32	207	16.1	236	1 CD81_HUMAN	P18582 homo sapien
33	206	16.1	225	1 CD9_RAT	P40241 rattus norv

RESULT 1									
T4S9_HUMAN									
ID	T4S9_HUMAN	STANDARD;	227	1	CD9_CERAE	P30409	cercopithec		
AC	O60628; O60746; Q9JLY1;		227	1	CD9_HUMAN	P21926	homo sapien		
DT	16-OCT-2001 (Rel. 40, Created)		281	1	CD37_MOUSE	Q61470	mus musculus		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		281	1	CD37_RAT	P31053	rattus norv		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		225	1	CD9_FELCA	P40239	felis silve		
DE	Transmembrane 4 superfamily, member 8 (Tetraspanin 5) (Tspan-5)		221	1	TSN2_MOUSE	O92236	mus musculus		
DE	(Tetraspanin NET-4).		221	1	TSN2_RAT	Q9JJW1	rattus norv		
GN	TM4SF9 OR TSPAN5.		218	1	IM23_SCHMA	Q26499	schistosoma		
OS	Homo sapiens (Human), and		218	1	IM23_SCHHA	Q26499	schistosoma		
OC	Mus musculus (Mouse).		225	1	CD9_BOVIN	P27591	bos taurus		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		218	1	IM23_SCHJA	O60636	homo sapien		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
OX	NCBI_TaxID=9606, 10090;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	SPECIES=Human;								
RX	MEDLINE=98390278; PubMed=9714763;								
RA	Todd S.C., Doctor V.S., Levy S.;								
RT	"Sequences and expression of six new members of the tetraspanin/TM4SF family.";								
RL	Biochim. Biophys. Acta 1399:101-104 (1998).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	SPECIES=Human;								
RA	Rubinstein E., Serru V., Boucheix C.;								
RT	"New tetraspans identified in the EST database.";								
RL	Submitted (MAY-1998) to the ENBL/GenBank/DBJ databases.								
RN	[3]								
RP	SEQUENCE FROM N.A.								
RC	SPECIES=Human; TISSUE=Pancreas;								
RX	MEDLINE=22388257; PubMed=12477932;								
RA	Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,								
RA	Altschuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,								
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,								
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,								
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,								
RA	Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,								
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,								
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,								
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,								
RA	Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,								
RA	Blakesley R.C., Grimwood J., Schmutz J., Myers R.M.,								
RA	Rodriguez A.C., Schmeissner M., Krzywicki M.I., Skalska U., Smalhus D.E.,								
RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;								
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";								
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).								
RN	[4]								
RP	SEQUENCE FROM N.A.								
RC	SPECIES=Mouse; STRAIN=Swiss Webster / NIH;								

us-09-972-970-4.isp

Mon Nov 24 10:01:53 2003

Garcia-Frigola C., de Lecea L., Soriano E.;  
"Mouse Tspan-5 cDNA cloning."  
Submitted (JAN-1999) to the EMBL/GenBank/DBSJ databases.  
-!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
-!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF053455; AAC69712.1; -  
EMBL; AF065389; AAC17120.1; -  
EMBL; BC009704; AAH09704.1; -  
EMBL; AF121344; AAF28869.1; -  
PIR; A59261; A59261.  
MGD; MGI:1928096; Tmsf9.  
GO; GO:0016021; C:integral to membrane; TAS.  
GO; GO:0008583; P:mystery cell fate differentiation (sensu Dr. .; TAS.  
InterPro; IPR000301; Transmem 4.  
Pfam; PF00335; Transmembrane4; 1.  
PRINTS; PR00259; TMFOUR.  
PROSITE; PS00421; TM4\_1; FALSE\_NEG.  
Glycoprotein; Transmembrane.  
CYTOPLASMIC (POTENTIAL).  
DOMAIN 1 17  
TRANSMEM 18 38  
DOMAIN 39 61  
TRANSMEM 62 82  
DOMAIN 83 92  
TRANSMEM 93 112  
DOMAIN 114 232  
TRANSMEM 233 253  
DOMAIN 254 268  
CARBOHYD 49 49  
CARBOHYD 169 169  
CARBOHYD 174 174  
CARBOHYD 232 232  
CONFLICT 91 94  
SEQUENCE 268 AA; 20327 MW; 7F4480ED0FA6192D CRC64;  
-----  
Query Match  
Best Local Similarity 79.9%; Score 885; DB 1; Length 268;  
Matches 155; Conservative 16; Mismatches 21; Indels 2; Gaps 1;  
-----  
QY 1 MFGKHQFQEPVEVCGCGKYFLFGFNVFWLGLAIGLWAWGKGVLSNISALTLGG 60  
DB 1 MSGK--HYKGPEVSCIKYFIFGFNVFWLGLAIGLWAWGKGVLSNISALTLGG 58  
-----  
QY 61 LDPVWLVVVGWVMSVLFAGCIGALRENTLKKFVSFLGIFLELATGILAFVKDW 120  
DB 59 FDPVWLVVVGWVMSVLFAGCIGALRENTLKKFVSFLGIFLELATGILAFVKDW 118  
-----  
QY 121 IRDLQNLFFNNVKNYRDDIDQLNLDFAQYWSCCGAGPDWNLNIFNCTDLPQRE 180  
DB 119 IKDQLVFFNNINRVRDRDIDQLNLDFTQYWCQCGAGPDWNLNIFNCTDLPQRE 178  
-----  
QY 181 RGVFPFSCVDRDPA 194  
DB 179 RGVFPFSCCTKDP 192  
-----  
RESULT 2  
TNE7 HUMAN STANDARD; PRT; 294 AA.  
ID TNE7 HUMAN  
AC O95858;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
Tetraspanin NET-7.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN SEQUENCE FROM N.A.  
RP Rubinstein E., Serru V., Dessen P., Boucheix C.;  
RA "New tetraspanins identified in the EST database."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBSJ databases.  
RN SEQUENCE FROM N.A.  
RP TISSUE=Kidney; PubMed=12477932;  
RX MEDLINE=22388257; Feingold E.A., Grouse L.H., Derge J.G.,  
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.N., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsien F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Touchman J.W., Green E.D., Dickson M.C.,  
RA Blakesley R.W., Grimwood J., Schmutz J., Myers R.M.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Skalska U., Smalius D.E.,  
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,  
RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002). (Probable).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF120266; AAD17295.1; -  
EMBL; BC003157; AAH04151.1; -  
EMBL; BC004161; AAH04161.1; -  
GO; GO:0005887; C:integral to plasma membrane; TAS.  
GO; GO:0005624; C:membrane fraction; TAS.  
InterPro; IPR000301; Transmem 4.  
Pfam; PF00335; Transmembrane4; 1.  
PROSITE; PS00421; TM4\_1; FALSE\_NEG.  
Glycoprotein; Transmembrane.  
CYTOPLASMIC (POTENTIAL).  
DOMAIN 1 23  
TRANSMEM 24 44  
DOMAIN 45 62  
TRANSMEM 63 83  
DOMAIN 84 93  
TRANSMEM 94 114  
DOMAIN 115 235  
TRANSMEM 236 256  
DOMAIN 257 294  
TRANSMEM 295 318  
CARBOHYD 118 118  
CARBOHYD 189 189  
CARBOHYD 230 230  
SEQUENCE 294 AA; 33165 MW; 71A6DC64D5CA6BAE CRC64;  
-----  
Query Match  
Best Local Similarity 33.9%; Score 324; DB 1; Length 294;  
Matches 64; Conservative 45; Mismatches 70; Indels 10; Gaps 3;  
-----  
QY 18 KYFLFGFNVFWLGLAIGLWAWGKGVLSNISALTLGGDPVWLVVVGWVMSVL 77

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:35:18 ; Search time 18 Seconds  
(without alignments)  
608.735 Million cell updates/sec

Title: US-09-972-970-4

Perfect score: 1283

Sequence: 1 MPKGKHQHFQBEVGCCKGYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	885	69.0	268	1 T4S9 HUMAN	O60628 homo sapien
2	324	25.3	294	1 TNE7 HUMAN	O95858 homo sapien
3	251	19.6	238	1 T4S7 MOUSE	O9dck3 mus musculu
4	243	18.9	238	1 T4S7 HUMAN	O14817 homo sapien
5	242	18.9	253	1 C151 RAT	O9qza6 rattus norv
6	240.5	18.7	245	1 T4S6 HUMAN	O43657 homo sapien
7	240	18.7	249	1 T4S2 HUMAN	P41732 homo sapien
8	239.5	18.7	245	1 T4S6 MOUSE	O70401 mus musculu
9	239	18.6	218	1 CD53 MOUSE	Q61451 mus musculu
10	239	18.6	253	1 C151 CERAE	Q9ymy2 cercopithe
11	239	18.6	253	1 C151 MOUSE	O35566 mus musculu
12	236	18.4	239	1 TNE5 HUMAN	O75954 homo sapien
13	236	18.4	249	1 T4S2 MOUSE	Q62283 mus musculu
14	235.5	18.4	219	1 CD55 HUMAN	P19397 homo sapien
15	234	18.2	253	1 C151 HUMAN	P48509 homo sapien
16	233.5	18.0	237	1 CD53 RAT	P24485 rattus norv
17	231	18.0	218	1 T4S3 HUMAN	P19075 homo sapien
18	228	17.8	267	1 CD82 HUMAN	P27701 homo sapien
19	220	17.1	237	1 CD63 MOUSE	P41731 mus musculu
20	217	16.9	266	1 CD82 MOUSE	P40237 mus musculu
21	215.5	16.8	236	1 CD63 BOVIN	O9x8k2 bos taurus
22	215.5	16.8	266	1 CD82 RAT	O70352 rattus norv
23	214.5	16.7	225	1 CD9 PIG	O8wmq3 sus scrofa
24	214	16.7	237	1 CD63 HUMAN	O89623 homo sapien
25	214	16.7	237	1 CD63 RABIT	O28709 cryptotagus
26	211	16.4	237	1 CD63 RAT	P28648 rattus norv
27	210	16.4	236	1 CD81 MOUSE	P35762 mus musculu
28	208	16.2	225	1 CD9 MOUSE	P40240 mus musculu
29	208	16.2	236	1 CD81 RAT	Q62745 rattus norv
30	208	16.2	236	1 CD81 SAGOE	Q9n019 saguinus oe
31	207	16.1	236	1 CD81 CERAE	O97703 cercopithe
32	207	16.1	236	1 CD81 HUMAN	P18582 homo sapien
33	206	16.1	225	1 CD9 RAT	P40241 rattus norv

34	202	15.7	227	1	CD9 CERAE	P30409 cerco
35	201	15.7	227	1	CD9 HUMAN	P21926 homo sa
36	200.5	15.6	281	1	CD37 MOUSE	O61470 mus muscu
37	200.5	15.6	281	1	CD37 RAT	P31053 rattus norv
38	199	15.5	225	1	CD9 FELCA	Q92239 felis silve
39	197	15.4	221	1	TSN2 MOUSE	O922j6 mus musculu
40	197	15.4	221	1	TSN2 RAT	Q91jw1 rattus norv
41	195.5	15.2	218	1	IM23 SCHMA	P19331 schistosoma
42	190.5	14.8	218	1	IM23 SCHHA	O26499 schistosoma
43	189	14.7	225	1	CD9 BOVIN	P30932 bos taurus
44	188.5	14.7	218	1	IM23 SCHUA	P27591 schistosoma
45	188	14.7	221	1	TSN2 HUMAN	O60636 homo sapien

## ALIGNMENTS

RESULT 1  
T4S9 HUMAN  
ID T4S9 HUMAN STANDARD; PRT; 268 AA.  
AC O60628, O60746; Q9JLY1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Transmembrane 4 superfamily, member 8 (Tetraspanin 5) (Tspan-5)  
DE (Tetraspanin NET-4).  
GN TM4SF9 OR TSPAN5.  
OS Homo sapiens (Human), and  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606, 10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX MEDLINE=98390278; PubMed=9714763;  
RA Todd S.C., Doctor V.S., Levy S.;  
RT "Sequences and expression of six new members of the tetraspanin/TM4SF family";  
RL Biochim. Biophys. Acta 1399:101-104 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX Rubinstein E., Serru V., Boucheix C.;  
RT "New tetraspans identified in the EST database";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human; TISSUE=Pancreas;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smaluk D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Mouse; STRAIN=Swiss Webster / NIH;

```

RA Garcia-Frigola C., de Lecea L., Soriano E.;
RT "Mouse Tspan-5 cDNA cloning.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF053455; AAC69712.1; -
DR EMBL; AF065389; AAC17120.1; -
DR EMBL; BC009704; AAH09704.1; -
DR EMBL; AF121344; AAF28869.1; -
DR PIR; A59261; A59261.
DR MGI; MGI:1928096; TM4sf9.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0008583; P:mystery cell fate differentiation (sensu Dr. . .); TAS.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
DR PROSITE; PS00421; TM4_1; FALSE_NEG.
KW Glycoprotein; Transmembrane.
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 18 38 POTENTIAL.
FT DOMAIN 39 61 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 62 82 POTENTIAL.
FT DOMAIN 83 92 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 93 113 POTENTIAL.
FT DOMAIN 114 232 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 233 253 POTENTIAL.
FT DOMAIN 254 268 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 49 49 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 91 94 MISSING (IN REF. 1).
SQ SEQUENCE 268 AA; 30337 MW; 7F4480B0FA6192D CRC64;

Query Match 69.0%; Score 885; DB 1; Length 268;
Best Local Similarity 79.9%; Pred. No. 1.4e-67;
Matches 155; Conservative 16; Mismatches 21; Indels 2; Gaps 1;

Qy 1 MPGRKHQFOEPEVCCGKYLFGFNIVFWVLGALFLAIGLWAWGKGVLSNISALTDLGG 60
Db 1 MSGK--HYKGEVSCCKYFIFFGNVIFWFLGITFLGLWAWWNEKGVLSNISITDLGG 58

Qy 61 LDPVWLVFWVGVMVFLGAGCIGALRENTFLKFFSVFLGLIFFLLELATGILAFVFKDW 120
Db 59 FDPVWLVFWVGVMVFLGAGCIGALRENTFLKFFSVFLGLIFFLLELATGILAFVFKDW 118

Qy 121 IRDLNFFINNKKAYRDDIDLQNLIDFAQYWCSCGAGNDNNLNYFNCTDLPNSRE 180
Db 119 IKQLYFFINNIRAYRDDIDLQNLIDFTQYWCSCGAGNDNNLNYFNCTDLPNSRE 178

Qy 181 RCGVPFSCCVKDP 194
Db 179 RCGVPFSCCTKDP 192

RESULT 2
TNE7 HUMAN STANDARD; PRT; 294 AA.
AC O95858;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tetraspanin NET-7.
GN NET7.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rubinstein E., Serru V., Desseix P., Boucheix C.;
RT "New tetraspans identified in the EST database.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF120266; AAD17295.1; -
DR EMBL; BC003157; AAH03157.1; -
DR EMBL; BC004161; AAH04161.1; -
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PROSITE; PS00421; TM4_1; FALSE_NEG.
KW Glycoprotein; Transmembrane.
FT DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 24 44 POTENTIAL.
FT DOMAIN 45 62 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 63 83 POTENTIAL.
FT DOMAIN 84 93 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 94 114 POTENTIAL.
FT DOMAIN 115 235 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 236 256 POTENTIAL.
FT DOMAIN 257 294 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 294 AA; 33165 MW; 71A6DC64D6CA6BAE CRC64;

Query Match 25.3%; Score 324; DB 1; Length 294;
Best Local Similarity 33.9%; Pred. No. 2.6e-20;
Matches 64; Conservative 45; Mismatches 70; Indels 10; Gaps 3;

Qy 18 KYFLFGFNIVFWVLGALFLAIGLWAWGKGVLSNISALTDGLDPLWLVFWVGVMVSL 77

```



```

Db 21 KFSLIITYSTVFWLIGALVLSVCIYAEVERQKYKTLIES-----AFLAPAILILLGVVPMV 76
Qy 78 GPAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDWIRDLQNFNNVKKAYR 137
Db 77 SFIGVLASRLNLYLQAFMYILGICLIMELIGGVVALTFRNQTFDLNDRIRGIENY 136
Qy 138 DDIDLQNLIDFAQEWSCGARGPNDWNLNIFNCTDLNPSRRCGVFPSCCVRDPAMSS 197
Db 137 DDLDFKNTIMDFVQKFKCGGDEYRDWSKNQYHDCS--APGPLACGVPTCCIR----NT 190
Qy 198 TPSVAMMSG 206
Db 191 TEVNTMTCG 199

RESULT 3
T4S7 MOUSE
ID T4S7 MOUSE STANDARD; PRT; 238 AA.
AC Q9DCK3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane 4 superfamily, member 7.
GN TM4SF7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalon D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length

```

```

RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBUNIT: FORMS A COMPLEX WITH INTEGRINS (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AK002709; BAB22301.1; -
CC EMBL; BC003482; AAH03482.1; -
CC MGD; MGI:1928097; Tm4sf7.
CC InterPro; IPR000301; Transmem 4.
CC Pfam; PF00335; transmembrane4; 1.
CC PRINTS; PR00259; TMFOUR.
CC PROSITE; PS00421; TM4_1; 1.
KW Glycoprotein; Transmembrane.
FT DOMAIN 1 13
FT TRANSMEM 14 34 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 55 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 56 76 POTENTIAL.
FT DOMAIN 77 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 106 POTENTIAL.
FT DOMAIN 107 201 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 202 222 POTENTIAL.
FT DOMAIN 223 238 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 238 AA; 26053 MW; AA916EF607877FA CRC64;

Query Match 19.6%; Score 251; DB 1; Length 238;
Best Local Similarity 31.3%; Pred. No. 3e-14;
Matches 67; Conservative 28; Mismatches 95; Indels 24; Gaps 9;

Qy 14 GCCG--KYFLPGFNIVFWLIGALFLATGLMANGKGVLSNLSALTDLGGLDPVWLVVVG 71
Db 4 GCLGVKYLMPAFNLLFWLGGCVLGVGLAATQGNFATLS--SSPFLSAANLLIVTG 61
Qy 72 GVMSVLGACIGALRENTFLKFFSVFLGLIFLELATGILAFVKDWIRDLQNFNN 131
Db 62 TFVNAIGFVGCIGALKENKCLLTFFVLLLVFLLEATIAVLFFAYSDKIDSYAQOOLK 121
Qy 132 NVKAY--RDDIDLQNLIDFAQEWSCGARGPNDWNLNIFNCTDLNPSRRCGVFPSCC 189
Db 122 GLHLYGTQGVNGLTNAWSIIQTDFRCGVSNYTDW-PEVY-----NATR----VPDSCC 170
Qy 190 VRDPAMSSTPSVAMSGSNWSWRAPYTPKAVW 223
Db 171 ----LEFSDSCGLHEPCTW-WKSPCVETVKA-W 197

RESULT 4
T4S7 HUMAN
ID T4S7 HUMAN STANDARD; PRT; 238 AA.
AC Q14817;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane 4 superfamily, member 7 (Novel antigen 2) (NAG-2)
DE (Tetraspanin 4) (Tspan-4).
GN TM4SF7 OR NAG2 OR TSPAN4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98030601; PubMed=9360996;

```



RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Baha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallation D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC SUMMARY:

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AF053453; AAC69710.1; --
DR EMBL; AF043906; AAC64257.1; --
DR EMBL; U84895; AAD00560.1; --
DR EMBL; AF134426; AAF08365.1; --
DR EMBL; AL035608; --; NOT ANNOTATED_CDS.
DR EMBL; BC012389; AAH12389.1; --
DR PIR; A59258; A59258.
DR Genew; HGNC:11858; TM4SF6.
DR MIM; 300191; --
DR InterPro; IPR000301; Transmem 4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
DR PROSITE; PS00421; TM4_1; 1.
KW Transmembrane; Glycoprotein; Polymorphism.
FT DOMAIN 1 19
FT TRANSMEM 20 40
FT FT DOMAIN 41 59
FT TRANSMEM 60 80
FT FT DOMAIN 81 93
FT TRANSMEM 94 114
FT DOMAIN 115 208
FT TRANSMEM 209 229
FT FT DOMAIN 230 245
FT CARBOHYD 134 134
FT VARIANT 108 108
FT FT SEQUENCE 245 AA; 27563 MW; 811FAB19C2805BE2 CRC64;
Query Match 18.7%; Score 240.5; DB 1; Length 245;
Best Local Similarity 31.6%; Pred.No.2.3e-13;
Matches 61; Conservative 30; Mismatches 83; Indels 19; Gaps 6;
QY 1 MFGKHQHFQEPVGVCGKYFLFGFNIVFWLGLFALGLWAGKGVLSNISALTDLGG 60
DB 1 MASPSRRLTQKPVITCFKSVLLIYTFIWTGVILLAVGIWG--KVSLENYFSLNNEKA 57
QY 61 LDPVWL FVVVGVMGVLFAGCIGALRNTFLFKPFSVFLGLIFPLELATGILAFVKDW 120
DB 58 TNVPFVLATGVVILLTGFCFATCRASMWMLKLYAMFLTVLFLVELVAAIVGVFRHE 117
QY 121 IRDQLNEFFNNNVKA--YRDDIDLQ-NLIDPAQEYNSCCGARGPDNWNLIYFNCITLN 176
DB 118 IR--NSFKNVYKALKQYNSG DYRSHAVDKIQNTLHCCGVTDYRDW-----TDTN 166
QY 177 PSRRCGVFPFSCC 189

```



Db 71 VFGLFCFATCRGSPWMLKLYAMFLSLVFLAEVLVAGISGFVFRHEIKOTFLRTYTDAMQT 130  
QY 136 YRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRRCRGVPPFSCVVD 192  
Db 131 YNGNDSRAVDHVRQSLSCGVQVNTWNTSTSPYF-----LEHGIPPPSCNMIS 178

RESULT 8  
T456\_MOUSE  
ID T456\_MOUSE STANDARD; PRT; 245 AA.  
AC O70401;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Transmembrane 4 superfamily, member 6 (Tetra-spanin 6) (Tspan-6).  
GN T456F6 OR TSPAN6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98390278; PubMed=9714763;  
RA Todd S.C., Doctor V.S., Levy S.;  
RT "Sequences and expression of six new members of the tetraspanin/TM4SF  
RT family";  
RL Biochim. Biophys. Acta 1399:101-104 (1998).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF053454; AAC69711.1; -  
CC PIR; A59260; A59260.  
CC MGI; 1926264; Tm4sf6.  
CC InterPro; IPR000301; Transmem 4.  
CC Pfam; PF00335; transmembrane4; 1.  
CC PRINTS; PR00259; TMFOUR.  
CC PROSITE; PS00421; TM4\_1; FALSE\_NEG.  
CC Transmembrane, Glycoprotein.  
FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 20 40 POTENTIAL.  
FT DOMAIN 41 59 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 60 80 POTENTIAL.  
FT DOMAIN 81 93 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 94 114 POTENTIAL.  
FT DOMAIN 115 208 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 209 229 POTENTIAL.  
FT DOMAIN 230 245 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 134 134 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 245 AA; 27333 MW; 25F68A1B3602A12 CRC64;

Query Match 18.7%; Score 239.5; DB 1; Length 245;  
Best Local Similarity 29.5%; Pred. No. 2.8e-13;  
Matches 56; Conservative 32; Mismatches 89; Indels 13; Gaps 3;

QY 1 MPGKHQHFQEPVGGCGVYFGFNIVFWVLGALFLAIGLWAGEKGVLSNLSALTDLGG 60  
Db 1 MASPSRRITQTPVITCLKSVLIYTFITWTVILLAVGIWG---KVSLENYFSLNLSKA 57  
QY 61 LDPVWLFWVVGVMVSLFAGCIGALRENTFLKFFSVFLGIFLELATGILAFVFKDW 120  
Db 58 TNVPFLVLTGTVILLGTGFCATCRISAMWKLKLYAMFLSLVFLAEVLVAGISGFVFRHE 117  
QY 121 IRDQINFFINNVAKYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSR 179

Db 118 IKNSFKSNYENALKKEYNSTGDIYSBAVDKIQSTLHCCGVTNYGDKWGTNYSET----- 171  
QY 180 ERGCVPPFSCC 189  
Db 172 ---GPPKSCC 178

RESULT 9  
CD53\_MOUSE  
ID CD53\_MOUSE STANDARD; PRT; 218 AA.  
AC Q61451; Q61721;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Leukocyte surface antigen CD53 (Cell surface glycoprotein CD53).  
GN CD53.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=129/SV;  
RX MEDLINE=93200067; PubMed=8452817;  
RA Wright M.D., Rochelle J.M., Tomlinson M.G., Seidlin M.F.,  
RA Williams A.F.;  
RT "Gene structure, chromosomal localization, and protein sequence of  
RT mouse CD53 (Cd53): evidence that the transmembrane 4 superfamily  
RT arose by gene duplication";  
RL Int. Immunol. 5:209-216 (1993).  
CC -!- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION IN HEMATOPOIETIC  
CC CELLS.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X97227; CAA65864.1; -  
CC EMBL; Z16078; CAA78892.1; -  
CC MGI; 88341; Cd53.  
CC InterPro; IPR000301; Transmem 4.  
CC Pfam; PF00335; transmembrane4; 1.  
CC PRINTS; PR00259; TMFOUR.  
CC PROSITE; PS00421; TM4\_1; 1.  
CC Glycoprotein; Antigen; Transmembrane.  
FT INIT MET 0 0 BY SIMILARITY.  
FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 11 31 POTENTIAL.  
FT DOMAIN 32 53 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 54 68 POTENTIAL.  
FT DOMAIN 69 79 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 80 105 POTENTIAL.  
FT DOMAIN 106 180 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 181 205 POTENTIAL.  
FT DOMAIN 206 218 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 118 118 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 128 128 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 147 147 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 218 AA; 24032 MW; 8CA592EADCE15E3D CRC64;

Query Match 18.6%; Score 239; DB 1; Length 218;  
Best Local Similarity 30.9%; Pred. No. 2.8e-13;  
Matches 63; Conservative 29; Mismatches 72; Indels 40; Gaps 7;  
QY 18 KYVLFQFNIVFWVLGALFLAIGLWAGEKGVLSNLSALTDLGGDPVWLVVVVGVM 74  
Db 9 KYVLFIFNLLFWVCGCCILGFGIYVQNTYGVLFRLNPLFT-LGN-----ILVVGSI 62

QY 75 SVLFGAGCIGALRENTFLKFFSVFLGLFFLELATGILAFVFKDWIRDLNFFNNVVK 134  
 Db 63 MVVAFGCGMSIKENKCLMSFFVLLIILAEVITAILLFYEQKJNTLVAEGLNDSIQ 122  
 QY 135 AYRDLIDQLNLDFAQEWSCCGARGPNDNINLYFNCTDNLNPSRRCGVFPSCCVRDPA 194  
 Db 123 VHSNDSNTMKAWDFIQTQLCCGVNGSSDWT-----SGPSPSC----- 160  
 QY 195 MSSTPSVAMSG-----SNWSWS 213  
 Db 161 -----ESGADVQGYNKAWSFHSN 180

## RESULT 10

C151\_CERAE STANDARD; PRT; 253 AA.  
 AC Q9MYM2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Platelet-endothelial tetraspan antigen 3 (PETA-3) (CD151 antigen).  
 GN CD151.  
 OS Cercopithecus aethiops (Green monkey) (Grivet), and  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopithecinae; Cercopithecus.  
 OX NCBI\_TaxID=9534, 9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shanmukhappa K., Kapil S.;  
 RT "CD151/PETA-3, a tetraspanin molecule, interacts with the 3'  
 RT untranslated region and partial nucleoprotein gene of porcine  
 RT reproductive and respiratory syndrome virus RNA.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; AF275665; AAF90151.1; -;  
 DR EMBL; AF275666; AAF90152.1; -;  
 DR InterPro; IPR000301; Transmem 4.  
 DR Pfam; PF00335; transmembrane4; 1.  
 DR PRINTS; PR00259; TMFOUR.  
 DR PROSITE; PS00421; TM4\_1; 1.  
 KW Glycoprotein; Transmembrane.  
 FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 19 39 POTENTIAL.  
 FT DOMAIN 40 57 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 58 78 POTENTIAL.  
 FT DOMAIN 79 91 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT DOMAIN 113 221 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 222 242 POTENTIAL.  
 FT DOMAIN 243 253 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 159 159 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 253 AA; 28438 MW; CAD2780B63F644A4 CRC64;

Query Match 18.6%; Score 239; DB 1; Length 253;  
 Best Local Similarity 31.3%; Pred. No. 3.2e-13;  
 Matches 61; Conservative 30; Mismatches 78; Indels 26; Gaps 7;

QY 8 FQEPVGC---CGKYFLFGNIVFWLGAFLALCLHWAGEKGVLSNISALTDLGGIDPV 64  
 Db 4 FNEKKTGTVCLLYLLFTYNCFFWLAGLWAVMGWITLALKS--DYISLASGYLATA 61

QY 65 WLFVVVGVMSVLGPGAGCIGALRENTFLKFFSVFLGLFFLELATGILAFVFKDWIRDO 124  
 Db 62 YILVWAGAVMVTVGLGCCCATFKERRNLLRYFILLIIFLEIITAGVLAYVY---YQO 117  
 QY 125 LNFFNNNNKAVRDDI-----DLQNLIDFAQEWSCCGARGPNDNINLYFNCTD 174  
 Db 118 LNTELKENLK---DTMAKRYHQPGHEAVTSAVDLQQLQEFHCSSNNNSQDWRDSEWIR--- 171  
 QY 175 LNPSRRCGVFPSCC 189  
 Db 172 LREARGRV-VPDSCC 185

## RESULT 11

C151\_MOUSE STANDARD; PRT; 253 AA.  
 AC O35566; O89118;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Platelet-endothelial tetraspan antigen 3 (PETA-3) (GP27) (Membrane  
 DE glycoprotein SFA-1) (CD151 antigen).  
 GN CD151 OR PETA3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97438231; PubMed=9294006;  
 RA Hasegawa H., Watanabe H., Nomura T., Utsumiya Y., Yanagisawa K.,  
 RA Fujita S.;  
 RT "Molecular cloning and expression of mouse homologue of SFA-1/PETA-3  
 RT (CD151), a member of the transmembrane 4 superfamily.";  
 RL Biochim. Biophys. Acta 1353:125-130(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=98267146; PubMed=9602068;  
 RA Fitter S., Seldin M.F., Ashman L.K.;  
 RT "Characterisation of the mouse homologue of CD151 (PETA-3/SFA-1);  
 RT genomic structure, chromosomal localisation and identification of 2  
 RT novel splice forms.";  
 RL Biochim. Biophys. Acta 1398:75-85(1998).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; D89290; BAA24447.1; -;  
 DR EMBL; AF033620; AAC25952.1; -;  
 DR EMBL; U89772; AAC25976.1; -;  
 DR MGI; MGI:1096360; Cd151.  
 DR InterPro; IPR000301; Transmem 4.  
 DR Pfam; PF00335; transmembrane4; 1.  
 DR PRINTS; PR00259; TMFOUR.  
 DR PROSITE; PS00421; TM4\_1; 1.  
 KW Glycoprotein; Transmembrane.  
 FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 19 39 POTENTIAL.  
 FT DOMAIN 40 57 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 58 78 POTENTIAL.  
 FT DOMAIN 79 91 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT DOMAIN 113 221 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 222 242 POTENTIAL.

FT DOMAIN 243 253 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 237 237 G -> S (IN REF. 1).  
 SQ SEQUENCE 253 AA; 28246 MW; AEBEEBCE2D765F1B CRC64;  
 Query Match 18.6%; Score 239; DB 1; Length 253;  
 Best Local Similarity 32.3%; Pred. No. 3.2e-13;  
 Matches 62; Conservative 28; Mismatches 82; Indels 20; Gaps 7;  
 QY 8 FOEPEVGC---CGKYFLFGFNIVFWLGAFLAIGLWAWGKGVLSNISALTDLGLDVPV 64  
 DB 4 FNEKATCGTVCCLKVLLFTYNCFWLAGLVAVGILWALKS--DYISLASSTYLATA 61  
 QY 65 WLFVVVGVMVSLGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDWIRDO 124  
 DB 62 YILVAGVVMVTVGLGCCATKERNLLRLYFILLIIFLEIITAGILAYVY----YQQ 117  
 QY 125 LNFFINNN-----VKAYRDD--IDLQNLIDFAQYWSCGARGPNDNLIYFNCTDLNP 177  
 DB 118 LNTLKENLKTQVMVKRYHQSGHEGVSSAVDKLQBFHCCGSGNSQDWDSEWIRSGEAD- 176  
 QY 178 SRRCGVFPSCC 189  
 DB 177 SRV---VPSDC 185

RESULT 12  
 TNE5\_HUMAN  
 ID TNE5\_HUMAN STANDARD; PRT; 239 AA.  
 AC O79554;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Tetraspan NET-5.  
 GN NET5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=2018533; PubMed=10719184;  
 RA Seru V., Dessen P., Boucheix C., Rubinstein E.;  
 RT "Sequence and expression of seven new tetraspans.";  
 RL Biochim. Biophys. Acta 1478:159-163(2000).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

-----  
 DR EMBL; AF089749; AAC35859.1;  
 DR GO; GO:0005887; C: integral to plasma membrane; TAS.  
 DR GO; GO:0005624; C: membrane fraction; TAS.  
 DR InterPro; IPR000301; Transmem. 4.  
 DR Pfam; PF00335; transmembrane4; 1.  
 DR PRINTS; PR00259; TMFOUR.  
 DR PROSITE; PS00421; TM4\_1; 1.  
 DR KW Glycoprotein; Transmembrane.  
 FT DOMAIN 1 13  
 FT TRANSMEM 14 34  
 FT POTENTIAL.  
 FT DOMAIN 35 55  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 56 76  
 FT POTENTIAL.  
 FT DOMAIN 77 85  
 FT CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 86 106  
 FT POTENTIAL.  
 FT DOMAIN 107 203  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 204 224  
 FT POTENTIAL.  
 FT DOMAIN 225 239  
 FT CYTOPLASMIC (POTENTIAL).

FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 239 AA; 26779 MW; DD7BA332BF6584EB CRC64;  
 Query Match 18.4%; Score 236; DB 1; Length 239;  
 Best Local Similarity 28.0%; Pred. No. 5.4e-13;  
 Matches 61; Conservative 30; Mismatches 91; Indels 36; Gaps 6;  
 QY 15 CCGKYFLFGFNIVFWLGAFLAIGLWAWGKGVLSNISALTDLGLDVPVWLVVVGVM 74  
 DB 7 CCLKYMFNLFNFWLGCGLGVLWISVSGNFATFS--PSFSLSANLVIAIGTIV 64  
 QY 75 SVLGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDWIRDOFFNNVK 134  
 DB 65 MYTGLGCLGAIKENKLLSFFVILLVILLAEILLILFFVYMDKVNNAKDLKEGLL 124  
 QY 135 AY--RDDIDLQNLIDFAQYWSCGARGPNDNLIYFNCTDLNPSRRCGVPFSCVDRD 192  
 DB 125 LYHTENNYGLKNAMNIIQAEMRCGVTDTYTDW-----YPLVGENTVDRCCM-- 171  
 QY 193 PAMSTSPVANMWSGNSWSRAPYTPKAVWASLRSGC 230  
 DB 172 -----ENSQCGRNATTP--LW---RTGC 190

RESULT 13  
 T4S2\_MOUSE  
 ID T4S2\_MOUSE STANDARD; PRT; 249 AA.  
 AC Q62283; O88429; Q9DBS3;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transmembrane 4 superfamily, member 2 (Cell surface glycoprotein A15)  
 DE (PE31) (TALLA homolog).  
 GN TM4SF2 OR MXS1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE OF 5-249 FROM N.A.  
 RC STRAIN=CS7BL/6; TISSUE=Spleen;  
 RC MEDLINE=20083609; PubMed=10617319;  
 RA Hosokawa Y., Ueyama E., Morikawa Y., Maeda Y., Seto M., Senba E.;  
 RT "Molecular cloning and expression of mouse PE31 (TALLA).";  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 5-249 FROM N.A.  
 RC STRAIN=CS7BL/6; TISSUE=Spleen;  
 RC MEDLINE=20083609; PubMed=10617319;  
 RA Hosokawa Y., Ueyama E., Morikawa Y., Maeda Y., Seto M., Senba E.;  
 RT "Molecular cloning of a cDNA encoding mouse A15, a member of the  
 RT transmembrane 4 superfamily, and its preferential expression in brain  
 RT neurons.";  
 RL Neurosci. Res. 35:281-290(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Lung;  
 RC MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,  
 RA Schirli L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Blake J., Offord D., Furuno M., Anon H., Baldarelli R., Barsh G.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., de Bonaldo M.F.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,



RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RT Hayashizaki Y.;  
RL "Functional annotation of a full-length mouse cDNA collection.";  
CC Nature 409:685-690(2001).  
CC -!- FUNCTION: MAY BE INVOLVED IN CELL PROLIFERATION AND CELL MOTILITY.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D26483; BAA05493.1; ALT INIT.  
CC EMBL; AF052492; AAC34579.1; ALT INIT.  
CC EMBL; AK004776; BAB23554.1; --  
CC MGD; MGI:1298407; Tm4sf2.  
CC InterPro: IPR000301; Transmem\_4.  
CC Pfam; PF00335; Transmembrane4; 1.  
CC PRINTS; PR00259; TMFOUR.  
CC PROSITE; PS00421; TM4\_1; 1.  
KW Glycoprotein; Transmembrane.  
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 17 40 POTENTIAL.  
FT DOMAIN 41 56 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 57 75 POTENTIAL.  
FT DOMAIN 76 86 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 87 112 POTENTIAL.  
FT DOMAIN 113 213 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 214 234 POTENTIAL.  
FT DOMAIN 235 249 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 155 155 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 158 158 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 146 146 R -> P (IN REF. 21).  
SQ SEQUENCE 249 AA; 27544 MW; 04B3A24D16B5DCD3 CRC64;  
  
Query Match 18.4%; Score 236; DB 1; Length 249;  
Best Local Similarity 28.8%; Pred. No. 5.7e-13;  
Matches 51; Conservative 35; Mismatches 79; Indels 12; Gaps 3;  
  
QY 16 CKYFLGFGNIVFWVLGALFLAIGLWANGKGVLSNISALTDGLDLPVWLVVVGVMYS 75  
DB 14 CLKTLTIYSFVFWITGVILLAVGVG---KLTGTYSLSIAENSTNAPYVLIGTGV 70  
QY 76 VLGFGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDWIRDLQNFNNVKA 135  
DB 71 VFGFGCATCRGSPWMLKLVAMFLSLVFLAELVAGISGFVRHEIKDTFRTYTDAMQN 130  
QY 136 YRDPIDLQNLIDFAQYWSCCGARGPDNWLNIYFNCTDLNPSRRCGVFPSCCVRD 192  
DB 131 YNGNDERSAVDHVQRSLSCGVQNYTNWSSPYF----LDH-----GIPSCCME 178  
  
RESULT 14  
ID CD53\_HUMAN STANDARD; PRT; 219 AA.  
AC P19397;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Leukocyte surface antigen CD53 (Cell surface glycoprotein CD53).  
GN CD53 OR MOX44.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=91055810; PubMed=1700763;  
RA Angelisova P., Vilec C., Stefanova I., Lipoldova M., Horejsi V.;  
RT "The human leukocyte surface antigen CD53 is a protein structurally  
RT similar to the CD37 and MRC OX-44 antigens.";  
RL Immunogenetics 32:281-285(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91079522; PubMed=2258620;  
RA Amiot M.;  
RT "Identification and analysis of cDNA clones encoding CD53. A pan-  
RT leukocyte antigen related to membrane transport proteins.";  
RL J. Immunol. 145:4322-4325(1990).  
CC -!- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION IN HEMATOPOIETIC  
CC CELLS.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: B CELLS, MONOCYTES, MACROPHAGES, NEUTROPHILS,  
CC SINGLE (CD4 OR CD8) POSITIVE THYMOCYTES, PERIPHERAL T CELLS.  
CC -!- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD53 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd53.htm".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M37033; AAA35663.1; --  
CC EMBL; M60871; AAA51951.1; --  
CC PIR; A37243; A37243.  
CC Genew; HGNC:1686; CD53.  
CC MIM; 151525; --  
CC InterPro: IPR000301; Transmem\_4.  
CC Pfam; PF00335; Transmembrane4; 1.  
CC PRINTS; PR00259; TMFOUR.  
CC PROSITE; PS00421; TM4\_1; 1.  
KW Glycoprotein; Antigen; Transmembrane.  
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 12 32 POTENTIAL.  
FT DOMAIN 33 54 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 55 69 POTENTIAL.  
FT DOMAIN 70 80 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 81 106 POTENTIAL.  
FT DOMAIN 107 181 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 182 206 POTENTIAL.  
FT DOMAIN 207 219 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 129 129 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 148 148 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 219 AA; 24341 MW; 48D4633DBC9110B6 CRC64;  
  
Query Match 18.4%; Score 235.5; DB 1; Length 219;  
Best Local Similarity 31.0%; Pred. No. 5.5e-13;  
Matches 54; Conservative 31; Mismatches 64; Indels 25; Gaps 5;  
  
QY 18 KYFLGFGNIVFWVLGALFLAIGLW--AWGKGVLSNISALTDGLDLPVWLVVVGVM 74  
DB 10 KYVLFNLLFWICGCCILGFIYLLIHNHNGVLPHNLPSLT-LGN-----VFVVGSI 63  
QY 75 SVLGFGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDWIRDLQNFNNVKA 134  
DB 64 MVVAFGLGMSGSIKENKCLMSFFILLIILAEVTLAILLVYRQKLNEYVAKGLTDSIH 123  
QY 135 AYRDDIDLQNLIDFAQYWSCCGARGPDNWLNIYFNCTDLNPSRRCGVFPSC 188  
DB 124 RYHSDNSTKAANDSIQSFLOCCGINTGSDWT-----SGPPASC 161  
  
RESULT 15  
ID C151\_HUMAN STANDARD; PRT; 253 AA.  
C151\_HUMAN

```

or send an email to license@lsb-sib.ch).
-----
CC      EMBL; U14650; AAA87064.1; -.
DR      EMBL; D29963; BAA06229.1; -.
DR      EMBL; AF315942; AAK14179.1; -.
DR      EMBL; BC001374; AAH01374.1; -.
DR      EMBL; BC013302; AAH13302.1; -.
DR      Genew; HGNC:1630; CD151.
DR      MIM; 602243; -.
DR      GO; GO:0005887; C:integral to plasma membrane; TAS.
DR      GO; GO:0005624; C:membrane fraction; TAS.
DR      InterPro; IPR000301; Transmem_4.
DR      Pfam; PF00335; transmembrane4; 1.
DR      PRINTS; PR00259; TMFOUR.
DR      PROSITE; PS00421; TM4 1; 1.
DR      KEGG; Glycoprotein; Transmembrane;
KW      DOMAIN 1 18
FT      TRANSMEM 19 39
FT      DOMAIN 40 57
FT      TRANSMEM 58 78
FT      DOMAIN 79 91
FT      TRANSMEM 92 112
FT      DOMAIN 113 221
FT      TRANSMEM 222 242
FT      DOMAIN 243 253
FT      CARBOHYD 159 159
FT      VARIANT 132 132
FT      VARIANT 137 137
FT      S -> P.
FT      /FTId=VAR_012490.
SQ      SEQUENCE 253 AA; 28113 MW; 5C81D7D2D750EAF CRC64;
                                         /FTId=VAR_012491.
Query Match 19.2%; Score 234; DB 1; Length 253;
Best Local Similarity 31.0%; Pred. No. 8.5e-13;
Matches 61; Conservative 29; Mismatches 77; Indels 30; Gaps 7;

Qy      8 FQPEVGC---CGKYFLFGNIVFWVLGALFLAIGLWAKGKGVLSNISALTDGLGDPV 64
Db      4 FNEKTTCTGVTVCLRYLLFTTNCFFLAGLAVMVGIIWTLAKS--DYISLASGYLATA 61
Qy      65 WLFVVVGVMVSLGFGACIGALRENTPLLKFFSFVLGLIIFLELATGILAFVKDWIRDQ 124
Db      62 YILVAVGTVMVTVGLGCCATFKERNLLRLYFILLIIFLEIIAGILAYAY----YQQ 117
Qy      125 LNPFINNKKAYRDDID-----LQNLIDPAQYWCSCARGPNQNLNIFYNCTD 174
Db      118 LNTLKENLK--DTWTRRYHQSHEAVTSAVDQLQOEPFHCCSGNSQDWRDSEWIR--- 171
Qy      175 LNPSRRCG--VPPSCC 189
Db      172 ---SQEAGGRVFPDSCC 185

Search completed: November 21, 2003, 13:44:05
Job time : 19 secs

```

Search completed: November 21, 2003, 13:44:05  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 17:14:19 ; Search time 148 Seconds  
(without alignments)  
7569.124 Million cell updates/sec

Title: US-09-972-970-2

Perfect score: 2538

Sequence: 1 ccacgcgtccggccagcc.....aaaaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A-COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B-COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTCUTS-COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	ID	Description
1	234	9.2	1932	4 US-09-482-273-41
2	95	3.7	399	3 US-09-188-930-62
3	95	3.7	399	3 US-09-188-930-243
4	95	3.7	399	4 US-09-312-283C-62
5	95	3.7	399	4 US-09-312-283C-243
6	67.8	2.7	1624	3 US-08-430-225A-19
7	58.4	2.3	473	4 US-09-702-705-1565
8	58.4	2.3	473	4 US-09-736-457-1565
9	57.8	2.3	1151	2 US-08-807-044-2
10	57.2	2.3	827	3 US-09-333-599-5
11	57.2	2.3	827	4 US-09-499-781-5
12	57.2	2.3	870	3 US-09-333-599-1
13	57.2	2.3	870	4 US-09-499-781-1
14	57.2	2.3	1344	3 US-08-705-771-8
15	54.8	2.2	977	2 US-08-855-140-2
16	54.8	2.2	977	4 US-09-016-434-938
17	54.2	2.1	1452	2 US-08-807-044-4
18	54.2	2.1	1452	5 PCT-US91-04986-1
19	52.2	2.1	687	1 US-08-254-493-2
20	52.2	2.1	687	1 US-08-408-222B-2
21	52.2	2.1	1120	1 US-08-254-493-3
22	52.2	2.1	1120	1 US-08-408-222B-3
23	46.8	1.8	560	3 US-09-221-298-12
24	46.2	1.8	371	4 US-09-643-587-292
25	46.2	1.8	371	4 US-09-480-884A-292
26	46.2	1.8	371	4 US-09-542-615A-292
27	46.2	1.8	371	4 US-09-606-421B-292

c	28	44.4	1.7	1637	3	US-08-852-824-3	Sequence 3, Appli
	29	43.6	1.7	390	3	US-09-197-649-7	Sequence 7, Appli
	30	43.2	1.7	639	4	US-09-252-991A-10430	Sequence 10430, A
	31	43.2	1.7	744	4	US-09-252-991A-10328	Sequence 10328, A
c	32	43.2	1.7	870	4	US-09-252-991A-10645	Sequence 10645, A
	33	43.2	1.7	1227	4	US-09-252-991A-10523	Sequence 10523, A
c	34	42.8	1.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	35	42.8	1.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	36	42.4	1.7	289	3	US-09-007-005-17	Sequence 17, Appli
	37	42.4	1.7	289	3	US-09-244-796-17	Sequence 17, Appli
c	38	42	1.7	801	3	US-09-020-956-16	Sequence 16, Appli
	39	42	1.7	801	3	US-09-030-607-16	Sequence 16, Appli
c	40	42	1.7	801	4	US-09-439-313-16	Sequence 16, Appli
	41	42	1.7	801	4	US-09-352-616A-16	Sequence 16, Appli
c	42	42	1.7	801	4	US-09-232-149A-16	Sequence 16, Appli
	43	41.4	1.6	7218	1	US-08-232-463-14	Sequence 14, Appli
	44	41	1.6	711	4	US-09-252-991A-5727	Sequence 5727, Ap
c	45	41	1.6	1548	4	US-09-252-991A-5713	Sequence 5713, Ap

## ALIGNMENTS

### RESULT 1

US-09-482-273-41  
; Sequence 41, Application US/09482273  
; Patent No. 6534631  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 71 Human Secreted Proteins  
; FILE REFERENCE: P2030P1  
; CURRENT APPLICATION NUMBER: US/09/482,273  
; CURRENT FILING DATE: 2000-01-13  
; EARLIER APPLICATION NUMBER: PCT/US99/15849  
; EARLIER FILING DATE: 1999-07-14  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/092,922  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/092,956  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 267  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 41:  
; LENGTH: 1932  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (2)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1022)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-482-273-41

Query Match 9.2%; Score 234; DB 4; Length 1932;

Best Local Similarity 58.8%; Pred. No. 1.4e-51;

Matches 448; Conservative 0; Mismatches 298; Indels 16; Gaps 2;

Qy	150	CCGAGGTCGGCTGCTGCGGGAATACCTCTGTTGGCTTCAACATGTTTCTGGTGC	209
Db	146	CTTTCGTGAGCCCGCTGGTGAATACCTGCTCTCTTCTTCAACATGCTCTTCTGGTGA	205
Qy	210	TGGGAGCCCTGCTCTGCGCCATCGGCTCTGGGCTGGGTGAGAGGGTGTCTCTCCA	269
Db	206	TTTCATGGTATGGTGGCTGTGGGTGTACGCTCGGCTAATGAAGCATGCAGAGCAG	265
Qy	270	ACATCTCTGCGCTGACCGATCTGGAGGCTCTGACCCCTGTGTGGCTGTTCTAGTGGTGTG	329
Db	266	CCCTAGCTGCTGGCAG-----TGGACCTGCCATCCTGCTGATCGTGGTGG	313



## RESULT 7

```
US-09-702-705-1565
; Sequence 1565, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1565
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (1)...(473)
; OTHER INFORMATION: n = A,T,C or G
US-09-702-705-1565

Query Match      2.3%; Score 58.4; DB 4; Length 473;
Best Local Similarity 50.6%; Pred. No. 9.3e-06;
Matches 137; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 321 TAGTGGTTGGAGCGTCATGTCGGTCTTCTCAGTGTCTTCTTGGCCCTCATCTTCTTCTCGAGC 380
Db 66 TCGCAGTGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 125

QY 381 AGAACACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 440
Db 126 AGAAGTATGTTTATGATCAGTTTGCCATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185

QY 441 TGGCAACAGGATCTTGGCTTCTTCAAGGATGATTCGAGACCGAGCTCAATTTCT 500
Db 186 TGGCCGAGCCATTCGCTATGTTNAGAGATAAGTGATGTCAGATTTATAACA 245

QY 501 TCATTAAACAACGTCGAGCTTCCGGATGACATTCGACCTCCAGAACCTCATTTGACT 560
Db 246 ACTTCGGCAGCAGATGGAGATTACCCGAAAAACAACACACTGNTTCNATCTCTGGACA 305

QY 561 TTGCTCAGGAATATTGGTCTTCTCGGAGC 591
Db 306 GGATCCAGCAGATTTTAAGTCTGTGGGC 336

RESULT 8
US-09-736-457-1565
; Sequence 1565, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
```

```
US-09-736-457-1565
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1565
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (1)...(473)
; OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1565

Query Match      2.3%; Score 58.4; DB 4; Length 473;
Best Local Similarity 50.6%; Pred. No. 9.3e-06;
Matches 137; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 321 TAGTGGTTGGAGCGTCATGTCGGTCTTCTCAGTGTCTTCTTGGCCCTCATCTTCTTCTCGAGC 380
Db 66 TCGCAGTGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 125

QY 381 AGAACACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 440
Db 126 AGAAGTATGTTTATGATCAGTTTGCCATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185

QY 441 TGGCAACAGGATCTTGGCTTCTTCAAGGATGATTCGAGACCGAGCTCAATTTCT 500
Db 186 TGGCCGAGCCATTCGCTATGTTNAGAGATAAGTGATGTCAGATTTATAACA 245

QY 501 TCATTAAACAACGTCGAGCTTCCGGATGACATTCGACCTCCAGAACCTCATTTGACT 560
Db 246 ACTTCGGCAGCAGATGGAGATTACCCGAAAAACAACACACTGNTTCNATCTCTGGACA 305

QY 561 TTGCTCAGGAATATTGGTCTTCTCGGAGC 591
Db 306 GGATCCAGCAGATTTTAAGTCTGTGGGC 336

RESULT 9
US-08-807-044-2
; Sequence 2, Application US/08807044
; Patent No. 5863735
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,044
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0224 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
```

```

; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNOT01
; CLONE: 663655
US-08-807-044-2

Query Match      2.3%; Score 57.8; DB 2; Length 1151;
Best Local Similarity 51.3%; Pred. No. 2.1e-05;
Matches 162; Conservative 0; Mismatches 148; Indels 6; Gaps 1;

QY 166 CGGAAATATCTCTGTTGGCTTCAACATTTTTCGTGGGTCGAGGAGCCCTGTTCTTCCGATCGG 225
DB 237 COTCAAGTACCTCATGTTGCGCTTCAACCTGCTNTTCTGGCTGGAGGCTGTGGCGTGT 296
QY 226 GSCCATCGGCTCTGGGCTGGGTCGAGAGGGTGTCTCTCCAACATCTCTGCCCTGAC 285
DB 297 GGTGTGCGCATCTGCTGGCGCCACACAGGGAGCTTCGCCACGSGTGC-----CTC 350
QY 286 CGATCTGGGAGGCTCGACCCCTGTGCTGTTTCTAGTGGTGTGGAGGCTCATGTCGT 345
DB 351 TTCTTCCGTCCTGTGCGTGCCCACTGCTCATCATCACCAGGCTTGTGCATGCG 410
QY 346 GCTGGCTTTGCGGCTGATCGGGCTCTCCGGAGAACACTTTTCTGCTCAAGTTTTT 405
DB 411 CATCGGCTTCTGGGCTGCTGGTGCCTCAAGGAGAACAAAGTGCCTCTGCTCACATT 470
QY 406 CTCAGTGTTCCTTGGCCCTCATCTTCTCTGGAGCTGGCAACAGGATCTTGGCCTTGT 465
DB 471 CTCTCTGCTGCTGCTGCTGTTTCTCTGAGGAGGAGTGTGCGGCTTCTTCTTCTCGC 530
QY 466 ATTCAGGACTGGATT 481
DB 531 CTACAGGACAAGATT 546

RESULT 10
US-09-333-599-5
; Sequence 5, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/333,599
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (57)..(815)
US-09-333-599-5

Query Match      2.3%; Score 57.2; DB 3; Length 827;
Best Local Similarity 48.3%; Pred. No. 2.5e-05;
Matches 195; Conservative 0; Mismatches 203; Indels 6; Gaps 1;

QY 115 CACCATCCGGGCAAGCACAGCAGCACTTCCAGGAACCCGAGGTGCGGTGCTGCGGGAATA 174
DB 50 CCCAGGATGGTGAGTTCAACGAGAGAGAACATGTGGCACCGTTTGCCTCAAGTA 109
QY 175 CTTCCTGTTTGGCTTCAACATTTTCTGGGTGCTGGAGCCCTGTTCCTGGCCATCGG 234
DB 110 CTGCTGTTTACCTACAAATTGCTGCTTCTGGCTGGCTGGCTGCTGATGGCAGTGGG 169
QY 235 CCTCTGGGCTTGGGTCGAGAGGGTGTCTCTCCAAACATCTCTGGGCTGACCGATCTGGG 294
DB 170 CATCTGGACGCTGGCCCTCAAGAGTG-----ACTACATCAGCCCTGTGGCTCAGGCAC 223
QY 295 AGGCTCGACCCCTGTGTGGCTGTTTGTAGTGGTGGAGGGGTCAATGTCGCTGGGCTT 354
DB 224 CTACCTGGCCACAGCCCTACATCTCTGTTGGTGGCGGCACTGTCTGTCATGTCAGTGGGT 283
QY 355 TGCCGGCTGCATCGGGCTCTCCGGGAGAACACTTTTCTGCTCAAGTTTCTCAGTGT 414
DB 284 CTTGGGCTGCTGCGCCACCTTCAAGGAGCGTCGGAACCTGCTGCGCTGACTTCTATCCT 343
QY 415 CTTGGGCTCATCTTCTTCTCTGGAGCTGGCAACAGGATCTTGGGCTTCTGTTTCAAGGA 474
DB 344 GCTCTCATCATCTTCTGCTGGAGATCATGCTGTGATCTCTGCTACGCCCTACTACCA 403
QY 475 CTGGATTCGAGACCAAGCTCAATTTCTTCAATAACAAACAGTCA 518
DB 404 GCAGCTGAACACAGGAGCTCAAGGAGAACCTGAAGGACACCATGA 447

RESULT 11
US-09-499-781-5
; Sequence 5, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/499,781
; CURRENT FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 09/333,599
; PRIOR FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (57)..(815)
US-09-499-781-5

Query Match      2.3%; Score 57.2; DB 4; Length 827;
Best Local Similarity 48.3%; Pred. No. 2.5e-05;
Matches 195; Conservative 0; Mismatches 203; Indels 6; Gaps 1;

QY 115 CACCATCCGGGCAAGCACAGCAGCACTTCCAGGAACCCGAGGTGCGGTGCTGCGGGAATA 174
DB 50 CCCAGGATGGTGAGTTCAACGAGAGAGAACATGTGGCACCGTTTGCCTCAAGTA 109
QY 175 CTTCCTGTTTGGCTTCAACATTTTCTGGGTGCTGGAGCCCTGTTCCTGGCCATCGG 234
DB 110 CTGCTGTTTACCTACAAATTGCTGCTTCTGGCTGGCTGGCTGCTGTCATGGCAGTGGG 169
QY 235 CCTCTGGGCTTGGGTCGAGAGGGTGTCTCTCCAAACATCTCTGCGCTGACCGATCTGGG 294
DB 170 CATCTGGACGCTGGCCCTCAAGAGTG-----ACTACATCAGCCCTGTGGCTCAGGCAC 223
QY 295 AGGCTCGACCCCTGTGTGGCTGTTTGTAGTGGTGGAGGGGTCAATGTCGCTGGGCTT 354
DB 224 CTACCTGGCCACAGCCCTACATCTCTGTTGGTGGCGGCACTGTCTGTCATGTCAGTGGGT 283
QY 355 TGCCGGCTGCATCGGGCTCTCCGGGAGAACACTTTTCTGCTCAAGTTTCTCAGTGT 414
```



Db 284 CTGGGCTGCTGGCCACCTTCAAGGAGCGTGGAACTGCTGGCGCTGACTTATCTCT 343  
QY 415 CTTGGCTCATCTTCTTCTGAGCTGGCAACAGGATCTTGGCTTCTGTTTCAAGGA 474  
Db 344 GCTCCTCATCATCTTCTGCTGGAGATCATCGTGGTATCTTCTGCTAGCGCTACTACCA 403  
QY 475 CTGGATTCGAGACCGAGCTCAATTTCTTCAATTACACACAGTCA 518  
Db 404 GCAGCTGAACACGGAGCTCAAGGAGAACTGAAGGACACCATGA 447

## RESULT 12

US-09-333-599-1  
; Sequence 1, Application US/09333599  
; Patent No. 6245898  
; GENERAL INFORMATION:  
; APPLICANT: Testa, Jacqueline E.  
; APPLICANT: Quigley, James P.  
; APPLICANT: Seandel, Marco  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS  
; FILE REFERENCE: SUNY  
; CURRENT APPLICATION NUMBER: US/09/333,599  
; CURRENT FILING DATE: 1999-06-15  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 870  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-333-599-1

Query Match 2.3%; Score 57.2; DB 3; Length 870;  
Best Local Similarity 48.3%; Pred. No. 2.6e-05;  
Matches 195; Conservative 0; Mismatches 203; Indels 6; Gaps 1;  
QY 115 CACCATCGCGGCAAGCAGCAGCAGCTTCCAGGAACCCGAGGTCGGTCTGCGGGAATA 174  
Db 94 CCCAGGATGGGTGAGTTCAACGAGAGAGAGACATGTGGACCGTTTGGCTCAAGTA 153  
QY 175 CTTCTGTTTGGCTTCAACATTTTCTGGGTGCTGGAGCCCTGTTCTCGGCCATCGG 234  
Db 154 CTTGCTGTTTACCTACATTTGCTGCTTCTGGCTGGCTGGCTGCTCATGGCAGTGG 213  
QY 235 CTTCTGGGCTGGGTGAGAGGGTCTTCTCCACATCTCTGGCTGACCGATCTGGG 294  
Db 214 CATCTGGAGCGTGGCCCTCAAGAGTG-----ACTACATCAGCCTGCTGGCCTCAGGCAC 267  
QY 295 AGGCCTCGACCCCTGTGTGGCTGTTTGTAGTGTGGAGCGTCATGTCCGTGCTGGGTT 354  
Db 268 CTACCTGGCCACAGCTACATCTGTGTGGTGGGCGCACTGTCGTATGTTGACTGGGT 327  
QY 355 TGCCGGCTGCATCGGGGCTCTCCGGAGAACACTTTCCTGTCACAGTTTCTCAGTGT 414  
Db 328 CTTGGGCTGCTGGCCACCTTCAAGGAGCGTCCGAACCTGCTGGCCTGTACTTCTATCT 387  
QY 415 CTTGGCTCATCTTCTTCTGAGCTGGCAACAGGATCTTGGCTTCTGTTTCTCAGTGT 474  
Db 388 GCTCCTCATCATCTTCTGCTGGAGATCATCGTGTATCTCTCGCTAGCGCTACTACCA 447  
QY 475 CTGGATTCGAGACCGAGCTCAATTTCTTCAATTACACACAGTCA 518  
Db 448 GCAGCTGAACACGGAGCTCAAGGAGAACTGAAGGACACCATGA 491

## RESULT 13

US-09-499-781-1  
; Sequence 1, Application US/09499781  
; Patent No. 6498014  
; GENERAL INFORMATION:  
; APPLICANT: Testa, Jacqueline E.  
; APPLICANT: Quigley, James P.

; APPLICANT: Seandel, Marco  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS  
; FILE REFERENCE: SUNY  
; CURRENT APPLICATION NUMBER: US/09/499,781  
; CURRENT FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 09/333,599  
; PRIOR FILING DATE: 1999-06-15  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 870  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-499-781-1

Query Match 2.3%; Score 57.2; DB 4; Length 870;  
Best Local Similarity 48.3%; Pred. No. 2.6e-05;  
Matches 195; Conservative 0; Mismatches 203; Indels 6; Gaps 1;  
QY 115 CACCATCGCGGCAAGCAGCAGCAGCTTCCAGGAACCCGAGGTCGGTCTGCTGGGGAATA 174  
Db 94 CCCAGGATGGGTGAGTTCAACGAGAGAGAGACATGTGGACCGTTTGGCTCAAGTA 153  
QY 175 CTTCTGTTTGGCTTCAACATTTTCTGGGTGCTGGAGCCCTGTTCTCGGCCATCGG 234  
Db 154 CTTGCTGTTTACCTACATTTGCTGCTTCTGGCTGGCTGGCTGCTCATGGCAGTGG 213  
QY 235 CTTCTGGGCTGGGTGAGAGGGTCTTCTCCACATCTCTGGCTGACCGATCTGGG 294  
Db 214 CATCTGGAGCGTGGCCCTCAAGAGTG-----ACTACATCAGCCTGCTGGCCTCAGGCAC 267  
QY 295 AGGCCTCGACCCCTGTGTGGCTGTTTGTAGTGTGGAGCGTCATGTCCGTGCTGGGTT 354  
Db 268 CTACCTGGCCACAGCTACATCTGTGTGGAGATCATCGTGTATCTCTCGCTAGCGCTACTACCA 327  
QY 355 TGCCGGCTGCATCGGGGCTCTCCGGAGAACACTTTCCTGTCACAGTTTCTCAGTGT 414  
Db 328 CTTGGGCTGCTGGCCACCTTCAAGGAGCGTCCGAACCTGCTGGCCTGTACTTCTATCT 387  
QY 415 CTTGGCTCATCTTCTTCTGAGCTGGCAACAGGATCTTGGCTTCTGTTTCTCAGTGT 474  
Db 388 GCTCCTCATCATCTTCTGCTGGAGATCATCGTGTATCTCTCGCTAGCGCTACTACCA 447  
QY 475 CTGGATTCGAGACCGAGCTCAATTTCTTCAATTACACACAGTCA 518  
Db 448 GCAGCTGAACACGGAGCTCAAGGAGAACTGAAGGACACCATGA 491

## RESULT 14

US-08-705-771-8  
; Sequence 8, Application US/08705771  
; Patent No. 6054289  
; GENERAL INFORMATION:  
; APPLICANT: Paul Moore, Reiner Gentz, Hongjin Ji,  
; APPLICANT: Jian Ni and Jing-Shan Hu  
; TITLE OF INVENTION: Human Genes, Sequences and  
; TITLE OF INVENTION: Expression Products  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:43:38 ; Search time 30 Seconds  
(without alignments)  
1417.879 Million cell updates/sec

Title: US-09-972-970-4  
Perfect score: 1283  
Sequence: 1 MPKGHQHFQBPVEVCCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	1283	100.0	233	10	US-09-972-970-4
2	1066	83.1	270	10	US-09-934-268-2
3	1066	83.1	270	15	US-10-103-196-16
4	1066	83.1	270	15	US-10-103-196-16
5	885	69.0	271	9	US-09-925-299-978
6	885	69.0	271	11	US-09-925-299-978
7	791	61.7	268	15	US-10-103-196-17
8	699	54.5	193	9	US-09-729-674-170
9	638.5	49.8	270	11	US-09-796-753-126
10	638.5	49.8	270	11	US-09-905-674-2
11	638.5	49.8	270	11	US-09-892-877-267
12	638.5	49.8	270	11	US-09-948-783-269
13	625.5	48.8	270	11	US-09-892-877-126
14	625.5	48.8	270	11	US-09-948-783-127
15	619.5	48.3	270	11	US-09-957-187-16

16	482.5	37.6	122	14	US-10-042-417-38	Sequence 38, Appl
17	429	33.4	120	15	US-10-106-698-6296	Sequence 6296, Ap
18	412.5	32.2	245	11	US-09-984-271-133	Sequence 131, App
19	411.5	32.1	209	11	US-09-957-187-79	Sequence 79, Appl
20	383	29.9	111	15	US-10-103-196-22	Sequence 22, Appl
21	325	25.3	254	10	US-09-934-268-4	Sequence 4, Appli
22	325	25.3	254	15	US-10-162-435-38	Sequence 38, Appl
23	324	25.3	294	11	US-09-946-374-123	Sequence 123, App
24	324	25.3	294	12	US-10-015-387A-123	Sequence 123, App
25	324	25.3	294	12	US-10-006-130A-123	Sequence 123, App
26	324	25.3	294	12	US-10-199-672-278	Sequence 123, App
27	324	25.3	294	12	US-10-006-172A-123	Sequence 123, App
28	324	25.3	294	12	US-10-187-749-278	Sequence 278, App
29	324	25.3	294	12	US-10-194-457-278	Sequence 278, App
30	324	25.3	294	12	US-10-184-642-278	Sequence 278, App
31	324	25.3	294	12	US-10-196-747-278	Sequence 278, App
32	324	25.3	294	12	US-10-015-392A-123	Sequence 123, App
33	324	25.3	294	12	US-10-017-253A-123	Sequence 123, App
34	324	25.3	294	12	US-10-173-689-278	Sequence 278, App
35	324	25.3	294	12	US-10-173-690-278	Sequence 278, App
36	324	25.3	294	12	US-10-173-691-278	Sequence 278, App
37	324	25.3	294	12	US-10-173-692-278	Sequence 278, App
38	324	25.3	294	12	US-10-173-694-278	Sequence 278, App
39	324	25.3	294	12	US-10-173-698-278	Sequence 278, App
40	324	25.3	294	12	US-10-173-699-278	Sequence 278, App
41	324	25.3	294	12	US-10-173-707-278	Sequence 278, App
42	324	25.3	294	12	US-10-174-569-278	Sequence 278, App
43	324	25.3	294	12	US-10-174-583-278	Sequence 278, App
44	324	25.3	294	12	US-10-174-587-278	Sequence 278, App
45	324	25.3	294	12	US-10-174-589-278	Sequence 278, App

ALIGNMENTS

RESULT 1  
US-09-972-970-4  
; Sequence 4, Application US/09972970  
; Patent No. US20020164693A1  
; GENERAL INFORMATION:  
; APPLICANT: Shi et al.  
; TITLE OF INVENTION: TM4SF Receptor Polynucleotides, Poypeptides, and Antibodies  
; FILE REFERENCE: PT05EP1  
; CURRENT APPLICATION NUMBER: US/09/972,970  
; CURRENT FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: PCT/US01/11130  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/195,336  
; PRIOR FILING DATE: 2000-04-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-972-970-4

Query Match	100.0%	Score 1283;	DB 10;	Length 233;
Best Local Similarity	100.0%	Pred. No. 9.7e-119;		
Matches 233;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MPKGHQHFQBPVEVCCGKYFPGNIVFVVLGALFLAIGLWAWGKGVLSNISALTDLGG	60	
Db	1	MPKGHQHFQBPVEVCCGKYFPGNIVFVVLGALFLAIGLWAWGKGVLSNISALTDLGG	60	
Qy	61	LDPVWLFWVVGWVSVLGFAGCGALRENTFLKPFVSFLGLIFFLAATGILAFVKDOW	120	
Db	61	LDPVWLFWVVGWVSVLGFAGCGALRENTFLKPFVSFLGLIFFLAATGILAFVKDOW	120	
Qy	121	IRDQLNFFINNNKAYRDDIDLQNLIDFAQYVSCCGARGPNDWNLNIFNCTDLNFSRE	180	
Db	121	IRDQLNFFINNNKAYRDDIDLQNLIDFAQYVSCCGARGPNDWNLNIFNCTDLNFSRE	180	

```
Qy 181 RCGVPFSCVDRPAMSSTPSVAMSGSNWSRBPYTPKAVWASLRSGCRTT 233
Db 181 RCGVPFSCVDRPAMSSTPSVAMSGSNWSRBPYTPKAVWASLRSGCRTT 233

RESULT 2
US-09-934-268-2
; Sequence 2, Application US/09934268
; Patent No. US20020172986A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; TITLE OF INVENTION: A NOVEL HUMAN TETRASPANIN FAMILY
; MEMBER AND USES THEREOF
; FILE REFERENCE: 10448-079001
; CURRENT APPLICATION NUMBER: US/09/934,268
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/226,612
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-934-268-2

Query Match 83.1%; Score 1066; DB 10; Length 270;
Best Local Similarity 99.5%; Pred. No. 3.2e-97;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPKGKHQHFQEPVEVCGCGKYFLFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGG 60
Db 1 MPKGKHQHFQEPVEVCGCGKYFLFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGG 60

Qy 61 LDPVWLFWVVGVMVSLGFCAGCIGALRENTFLKFFSVFLGLIFPFLATGILAFVFKDW 120
Db 61 LDPVWLFWVVGVMVSLGFCAGCIGALRENTFLKFFSVFLGLIFPFLATGILAFVFKDW 120

Qy 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQBYNSCCGARGNDWNLNIYFNCTDLPNPSRE 180
Db 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQBYNSCCGARGNDWNLNIYFNCTDLPNPSRE 180

Qy 181 RCGVPFSCVDRPDA 194
Db 181 RCGVPFSCVDRPDA 194

RESULT 3
US-10-103-196-16
; Sequence 16, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT008P1
; CURRENT APPLICATION NUMBER: US/10/103,196
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/707,936
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/13504
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/178,770
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/149,447
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/138,573
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/137,797
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/135,122
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 16
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-196-16

Query Match 83.1%; Score 1066; DB 15; Length 270;
Best Local Similarity 99.5%; Pred. No. 3.2e-97;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPKGKHQHFQEPVEVCGCGKYFLFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGG 60
Db 1 MPKGKHQHFQEPVEVCGCGKYFLFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGG 60

Qy 61 LDPVWLFWVVGVMVSLGFCAGCIGALRENTFLKFFSVFLGLIFPFLATGILAFVFKDW 120
Db 61 LDPVWLFWVVGVMVSLGFCAGCIGALRENTFLKFFSVFLGLIFPFLATGILAFVFKDW 120

Qy 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQBYNSCCGARGNDWNLNIYFNCTDLPNPSRE 180
Db 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQBYNSCCGARGNDWNLNIYFNCTDLPNPSRE 180

Qy 181 RCGVPFSCVDRPDA 194
Db 181 RCGVPFSCVDRPDA 194

RESULT 4
US-10-162-435-36
; Sequence 36, Application US/10162435
; Publication No. US20030096305A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Leiby, Kevin R.
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND
; FILE REFERENCE: 10448-189001
; CURRENT APPLICATION NUMBER: US/10/162,435
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/836,499
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: PCT/US01/12420
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 60/197,507
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/891,008
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: PCT/US01/19963
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/214,220
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/860,868
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/16013
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,674
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/886,429
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: PCT/US01/20055
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/213,963
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 10/041,406
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: PCT/US02/00275
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/260,286
; PRIOR FILING DATE: 2001-01-08
```

; PRIOR APPLICATION NUMBER: US 09/934,268
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/41811
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,612
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-162-435-36

Query Match 83.1%; Score 1066; DB 15; Length 270;
Best Local Similarity 99.5%; Pred. No. 3.2e-97;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MPGKHQHFQEPVEVCGCGKYFLFGFNVFVWLVGALFLAIGLWAWGKGVLSNISALTDLGG 60
Db 1 MPGKHQHFQEPVEVCGCGKYFLFGFNVFVWLVGALFLAIGLWAWGKGVLSNISALTDLGG 60

Qy 61 LDPVWLFVWVGWMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW 120
Db 61 LDPVWLFVWVGWMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW 120

Qy 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRE 180
Db 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRE 180

Qy 181 RCGVPFSCCVRDPA 194
Db 181 RCGVPFSCCVRDPA 194

RESULT 5
US-09-925-299-978
; Sequence 978, Application US/09925299
; Publication No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 978
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-978

Query Match 69.0%; Score 885; DB 9; Length 271;
Best Local Similarity 79.9%; Pred. No. 2.5e-79;
Matches 155; Conservative 16; Mismatches 21; Indels 2; Gaps 1;

Qy 1 MPGKHQHFQEPVEVCGCGKYFLFGFNVFVWLVGALFLAIGLWAWGKGVLSNISALTDLGG 60
Db 4 MSGK--HYKGEVSCCIKYFIYFGFNVFWFLGITFLGIGLWAWNEKGVLSNISITDLGG 61

Qy 61 LDPVWLFVWVGWMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW 120
Db 62 FDPVWLFVWVGWMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW 121

Qy 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRE 180
Db 122 IKDQLYFFINNIRAYRDDIDLQNLIDFTQYEWQCCGAFGADDNLIYFNCTDSNASRE 181

Qy 181 RCGVPFSCCVRDPA 194
Db 182 RCGVPFSCCVRDPA 195

RESULT 6
US-09-925-299-978
; Sequence 978, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 978
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-978

Query Match 69.0%; Score 885; DB 11; Length 271;
Best Local Similarity 79.9%; Pred. No. 2.5e-79;
Matches 155; Conservative 16; Mismatches 21; Indels 2; Gaps 1;

Qy 1 MPGKHQHFQEPVEVCGCGKYFLFGFNVFVWLVGALFLAIGLWAWGKGVLSNISALTDLGG 60
Db 4 MSGK--HYKGEVSCCIKYFIYFGFNVFWFLGITFLGIGLWAWNEKGVLSNISITDLGG 61

Qy 61 LDPVWLFVWVGWMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW 120
Db 62 FDPVWLFVWVGWMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW 121

Qy 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRE 180
Db 122 IKDQLYFFINNIRAYRDDIDLQNLIDFTQYEWQCCGAFGADDNLIYFNCTDSNASRE 181

Qy 181 RCGVPFSCCVRDPA 194
Db 182 RCGVPFSCCVRDPA 195

RESULT 7
US-10-103-196-17
; Sequence 17, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT008P1
; CURRENT APPLICATION NUMBER: US/10/103,196
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/707,936
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/13504
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/178,770
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/149,447
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/138,573
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/137,797
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/135,122
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 44



```

; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-126

Query Match          49.8%; Score 638.5; DB 11; Length 270;
Best Local Similarity 49.4%; Pred. No. 5.5e-55;
Matches 119; Conservative 37; Mismatches 66; Indels 19; Gaps 2;

QY  5 HQHFOPEVGCCKYFLGFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGLDPV 64
Db   3 YRYSNAKSCWKYLLFSYNIIFWLAGVYFLGVGLWAMEKGVLSDLTKVTRMHGIDPV 62
QY  65 WLFVVVGVGMSVLGFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWIRDO 124
Db   63 VLVLWGVVMFTLGFAGCVGALRENTCLNFFCGTIVLFFLELAVAVLAFQDWRDR 122
QY  125 LNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGENDNNIYFNCTDLNPSRRCGV 184
Db   123 PREFFESNIKSYRDDIDLQNLIDSLQKQCCGAYGPDWDLNVYFNCGASYSREKCGV 182
QY  185 PFSCCVRDPAMSTPS-----VAMMSGSNW-----SWSSRAPYTPKAVWAS 225
Db   183 PFSCCVPDPAQKVNTQCGYDVRIOQLKSKWDESIPTKGCIOALESWLPRIYIVAGVFIA 242
QY  226 L 226
Db   243 I 243

RESULT 10
US-09-905-674-2
; Sequence 2, Application US/09905674
; Publication No. US2003039647A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Garcia, Pablo
; TITLE OF INVENTION: TETRASPAN PROTEIN AND USES THEREOF
; FILE REFERENCE: PP-01700.002/200130.521
; CURRENT APPLICATION NUMBER: US/09/905,674
; CURRENT FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-674-2

Query Match          49.8%; Score 638.5; DB 11; Length 270;
Best Local Similarity 49.4%; Pred. No. 5.5e-55;
Matches 119; Conservative 37; Mismatches 66; Indels 19; Gaps 2;

QY  5 HQHFOPEVGCCKYFLGFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGLDPV 64
Db   3 YRYSNAKSCWKYLLFSYNIIFWLAGVYFLGVGLWAMEKGVLSDLTKVTRMHGIDPV 62
QY  65 WLFVVVGVGMSVLGFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWIRDO 124
Db   63 VLVLWGVVMFTLGFAGCVGALRENTCLNFFCGTIVLFFLELAVAVLAFQDWRDR 122
QY  125 LNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGENDNNIYFNCTDLNPSRRCGV 184
Db   123 PREFFESNIKSYRDDIDLQNLIDSLQKQCCGAYGPDWDLNVYFNCGASYSREKCGV 182
QY  185 PFSCCVRDPAMSTPS-----VAMMSGSNW-----SWSSRAPYTPKAVWAS 225
Db   183 PFSCCVPDPAQKVNTQCGYDVRIOQLKSKWDESIPTKGCIOALESWLPRIYIVAGVFIA 242
QY  226 L 226
Db   243 I 243

; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-126

Query Match          49.8%; Score 638.5; DB 11; Length 270;
Best Local Similarity 49.4%; Pred. No. 5.5e-55;
Matches 119; Conservative 37; Mismatches 66; Indels 19; Gaps 2;

QY  5 HQHFOPEVGCCKYFLGFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGLDPV 64
Db   3 YRYSNAKSCWKYLLFSYNIIFWLAGVYFLGVGLWAMEKGVLSDLTKVTRMHGIDPV 62
QY  65 WLFVVVGVGMSVLGFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWIRDO 124
Db   63 VLVLWGVVMFTLGFAGCVGALRENTCLNFFCGTIVLFFLELAVAVLAFQDWRDR 122
QY  125 LNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGENDNNIYFNCTDLNPSRRCGV 184
Db   123 PREFFESNIKSYRDDIDLQNLIDSLQKQCCGAYGPDWDLNVYFNCGASYSREKCGV 182
QY  185 PFSCCVRDPAMSTPS-----VAMMSGSNW-----SWSSRAPYTPKAVWAS 225
Db   183 PFSCCVPDPAQKVNTQCGYDVRIOQLKSKWDESIPTKGCIOALESWLPRIYIVAGVFIA 242
QY  226 L 226
Db   243 I 243
```

```

RESULT 11
US-09-892-877-267
; Sequence 267, Application US/09892877
; Publication No. US2003007809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 267
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-892-877-267

Query Match          49.8%; Score 638.5; DB 11; Length 270;
Best Local Similarity 49.4%; Pred. No. 5.5e-55;
Matches 119; Conservative 37; Mismatches 66; Indels 19; Gaps 2;

QY  5 HQHFOPEVGCCKYFLGFGNIVFWVLGALFLAIGLWAMEKGVLSNISALTDLGLDPV 64
Db   3 YRYSNAKSCWKYLLFSYNIIFWLAGVYFLGVGLWAMEKGVLSDLTKVTRMHGIDPV 62
QY  65 WLFVVVGVGMSVLGFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWIRDO 124
Db   63 VLVLWGVVMFTLGFAGCVGALRENTCLNFFCGTIVLFFLELAVAVLAFQDWRDR 122
QY  125 LNFFINNKKAYRDDIDLQNLIDFAQYWSCCGARGENDNNIYFNCTDLNPSRRCGV 184
Db   123 PREFFESNIKSYRDDIDLQNLIDSLQKQCCGAYGPDWDLNVYFNCGASYSREKCGV 182
QY  185 PFSCCVRDPAMSTPS-----VAMMSGSNW-----SWSSRAPYTPKAVWAS 225
Db   183 PFSCCVPDPAQKVNTQCGYDVRIOQLKSKWDESIPTKGCIOALESWLPRIYIVAGVFIA 242
QY  226 L 226
Db   243 I 243

RESULT 12
US-09-948-783-269
; Sequence 269, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
```



Query Match	48.8%	Score 625.5;	DB 11;	Length 270;
Best Local Similarity	49.0%;	Pred. No. 1.1e-53;		
Matches 118;	Conservative 37;	Mismatches 67;	Indels 19;	Gaps 2
Qy	5	HQHFQEPVGGCGKYFLFGFNIVFWLCAFLAIGLWANGKGVLSNISALTDLGGLDPV	64	
Db	3	YYRYSNAKUSCWYKYLFFSYNIIIFLAGVFLGVGLWAMSEKGVLSDLTKVTRMHGIDPV	62	
Qy	65	WLFPVVGVMVSLVGFAGCIGIGALRENTFLFKFPFSVFLGLTFPLELATGILAFVFKWIRDO	124	
Db	63	VLVMVGVMFTLGFAGCVGALRENICLLNFFCGTIVLIFPLELAVAVLAFQDQVDR	122	
Qy	125	LNFFINNWKAYRDDIDLQNLIDFAQETWSCGARGPNDWNINIFNCTDNLNPSRRCGV	184	
Db	123	FRFPFENIKSYRDDIDLQNLIDSLQKACQCGAYGPDWDUNVYFNCSGASYSREKCGV	182	
Qy	185	PFSCCVDRPAMSSPTPS-----VAMMSGSNW-----SWSSRAPYTPKAVMAS	225	
Db	183	PFSCCVDPQAQKVNTQCGYDVRQLQSKWDESIPTKCIQALSWLPNNIYIVAGVFIA	242	
Qy	226	L 226		
Db	243	I 243		

```

RESULT 14
US-09-948-783-127
; Sequence 127, Application US/09948783
; Publication No. US2003010051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: PZ028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,923
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,925
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,928
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,920
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 127

```

```

, PRIOR FILING DATE: 1998-05-12
, PRIOR APPLICATION NUMBER: 60/085,180
, PRIOR FILING DATE: 1998-05-12
, PRIOR APPLICATION NUMBER: 60/085,927
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,906
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,924
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,922
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,921
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,923
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,925
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,928
, PRIOR FILING DATE: 1998-05-18
, PRIOR APPLICATION NUMBER: 60/085,920
, PRIOR FILING DATE: 1998-05-18
, NUMBER OF SEQ ID NOS: 465
, SOFTWARE: PatentIn Ver. 2.0
, SEQ ID NO 127

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 17:20:09 ; Search time 770 Seconds

(without alignment)  
10772.341 Mallion cell updates/sec

Title: US-09-972-970-2

Perfect score: 2538

Sequence: 1 ccacgcctccggccgcagc.....aaaaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2538	100.0	2538	10	US-09-972-970-2
2	1061.6	41.8	2461	12	US-10-252-157-153
3	790.4	31.1	3184	10	US-09-934-268-1
4	790.4	31.1	3184	14	US-10-162-435-35
5	787.2	31.0	1126	14	US-10-103-196-4
6	710.4	28.0	813	10	US-09-934-268-3
7	710.4	28.0	813	14	US-10-162-435-37
8	441.6	17.4	1655	9	US-09-925-299-205
9	441.6	17.4	1655	11	US-09-925-299-205
10	418.6	16.5	1174	14	US-10-103-196-5
11	402.2	15.8	1178	14	US-10-103-196-1
12	390.8	15.4	1988	11	US-09-822-846-126
13	357.4	14.1	2672	11	US-09-892-877-20
14	357.4	14.1	2672	11	US-09-948-783-20
15	353.4	13.9	2715	11	US-09-796-753-125
16	351.8	13.9	1388	11	US-09-905-674-1

17	343.8	13.5	1428	11	US-09-991-053-15	Sequence 15, Appl
18	343.8	13.5	1428	11	US-09-957-187-15	Sequence 15, Appl
19	333.6	13.1	1110	9	US-09-723-674-169	Sequence 169, App
20	321	12.6	1667	11	US-09-892-877-266	Sequence 266, App
21	321	12.6	1667	11	US-09-948-783-268	Sequence 268, App
22	295.8	11.7	864	10	US-09-875-440-1	Sequence 1, Appl
23	294.2	11.6	368	13	US-10-042-417-37	Sequence 37, Appl
24	254.4	10.0	816	11	US-09-991-053-78	Sequence 78, Appl
25	254.4	10.0	816	11	US-09-957-187-78	Sequence 78, Appl
c	240	9.5	839	14	US-10-106-698-2019	Sequence 2019, Ap
27	234	9.2	1362	11	US-09-984-271-41	Sequence 41, Appl
28	231.6	9.1	361	11	US-09-918-995-37420	Sequence 37420, A
29	230	9.1	411	11	US-09-918-995-7371	Sequence 7371, Ap
30	200.2	7.9	360	11	US-09-918-995-37648	Sequence 37648, A
31	152.4	6.0	476	11	US-09-918-995-31988	Sequence 31988, A
32	130.8	5.2	476	11	US-09-918-995-36537	Sequence 36537, A
33	128	5.0	1725	14	US-10-103-196-7	Sequence 7, Appl
34	127	5.0	1772	11	US-09-813-153-58	Sequence 58, Appl
35	126.2	5.0	1778	11	US-09-946-374-122	Sequence 122, App
36	126.2	5.0	1778	12	US-10-015-387A-122	Sequence 122, App
37	126.2	5.0	1778	12	US-10-006-130A-122	Sequence 122, App
38	126.2	5.0	1778	12	US-10-199-672-277	Sequence 277, App
39	126.2	5.0	1778	12	US-10-006-172A-122	Sequence 122, App
40	126.2	5.0	1778	12	US-10-187-749-277	Sequence 277, App
41	126.2	5.0	1778	12	US-10-194-457-277	Sequence 277, App
42	126.2	5.0	1778	12	US-10-184-642-277	Sequence 277, App
43	126.2	5.0	1778	12	US-10-196-747-277	Sequence 277, App
44	126.2	5.0	1778	12	US-10-015-392A-122	Sequence 122, App
45	126.2	5.0	1778	12	US-10-017-253A-122	Sequence 122, App

#### ALIGNMENTS

#### RESULT 1

US-09-972-970-2  
; Sequence 2, Application US/09972970  
; Patent No. US20020164693A1  
; GENERAL INFORMATION:  
; APPLICANT: Shi et al.  
; TITLE OF INVENTION: TMSF Receptor Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PT056P1  
; CURRENT APPLICATION NUMBER: US/09/972,970  
; CURRENT FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: PCT/US01/11130  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/195,336  
; PRIOR FILING DATE: 2000-04-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2538  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-972-970-2

Query Match	100.0%	Score 2538;	DB 10;	Length 2538;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2538;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CCACGCGTCCGGCCGCGAGCCGCGGCTAGCCCGGGCGGCTCTAGCCCGAGGGCGGCC 60		
Db	1	CCACGCGTCCGGCCGCGAGCCGCGGCTAGCCCGGGCGGCTCTAGCCCGAGGGCGGCC 60		
Qy	61	GTGAGGCGCGCATCCCGGGCCCGGCTCCGGTTCCCGGGCCCGGGCGGCTGCTCACCAT 120		
Db	61	GTGAGGCGCGCATCCCGGGCCCGGCTCCGGTTCCCGGGCCCGGGCGGCTGCTCACCAT 120		
Qy	121	GCGGGGCAAGCACACGACATTCACAGAACCCGAGGTCCGCTGCTGCGGGAATACTTCCT 180		
Db	121	GCGGGGCAAGCACACGACATTCACAGAACCCGAGGTCCGCTGCTGCGGGAATACTTCCT 180		
Qy	181	GTTGGCTTCAACATTGTTTCTGGGTCTGGGAGCCCTGTTCCTGGCCATCGGCCTCTG 240		

181 GTTGGCTTCAACATTTGTTTCTGGGTCTGGAGCCCTGTTCCTTGGCCATCGGCCTCTG 240  
241 GGCCTGGGTGAGAGGGTGTTCCTCCAAACATCTCTCGCTGACCCGATCTGGGAGGCCT 300  
241 GGCCTGGGTGAGAGGGTGTTCCTCCAAACATCTCTCGCTGACCCGATCTGGGAGGCCT 300  
301 CGACCTGTGTGGCTGTTTGTAGTGGTGGAGGCGTCATGTCGCTGGGTGTCGCGG 360  
301 CGACCTGTGTGGCTGTTTGTAGTGGTGGAGGCGTCATGTCGCTGGGTGTCGCGG 360  
361 CTGCATCGGGCTCTCCGGGAAACACTTTCCTGCTCAAGTTTTCTCAGTGTTCCTTGG 420  
361 CTGCATCGGGCTCTCCGGGAAACACTTTCCTGCTCAAGTTTTCTCAGTGTTCCTTGG 420  
421 CCTCATCTTCTCTGGAGCTGGCAACAGGATCTTGGCCCTTCGTATTTCAAGGACTGGAT 480  
421 CCTCATCTTCTCTGGAGCTGGCAACAGGATCTTGGCCCTTCGTATTTCAAGGACTGGAT 480  
481 TCGAGACAGCTCAATTTCTTCAATTAACAACAACGTCAAGGCTATCGGGATGACATTGA 540  
481 TCGAGACAGCTCAATTTCTTCAATTTAAACAACAACGTCAAGGCTATCGGGATGACATTGA 540  
541 CCTCAGAACTCATTTGCTTCAGGAATATTGGTCTTGTGCGGAGCCCGAGGCCC 600  
541 CCTCAGAACTCATTTGCTTCAGGAATATTGGTCTTGTGCGGAGCCCGAGGCCC 600  
601 TAATGACTGGAACTCAATATCTATTTCAACTGCACTGATGAACCCGAGCCGAGAGCG 660  
601 TAATGACTGGAACTCAATATCTATTTCAACTGCACTGATGAACCCGAGCCGAGAGCG 660  
661 CTGGGGGTGCCCTTCTCTGCTGTGTCAGGAGCCCTCGCATGTCTCAACACCCAGTGT 720  
661 CTGGGGGTGCCCTTCTCTGCTGTGTCAGGAGCCCTCGCATGTCTCAACACCCAGTGT 720  
721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGGCTCCATACACCAAAAGGC 780  
721 GGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGGCTCCATACACCAAAAGGC 780  
781 TGTGTGGCCAGTTTGAGAGTGTCTGAGAGCTGAGAGCTGATGATGATGATGATGATGAT 840  
781 TGTGTGGCCAGTTTGAGAGTGTCTGAGAGCTGAGAGCTGATGATGATGATGATGATGAT 840  
841 GTGGGCTATGCTCTCTCCAGATCTTTGGTATCTGCTGCGCCAGAACTTTGTGAGTGAC 900  
841 GTGGGCTATGCTCTCTCCAGATCTTTGGTATCTGCTGCGCCAGAACTTTGTGAGTGAC 900  
901 ATCAAGGAGTGAAGGCCCACTGGATCAAACTGATGATGATGATGATGATGATGATGAT 960  
901 ATCAAGGAGTGAAGGCCCACTGGATCAAACTGATGATGATGATGATGATGATGATGAT 960  
961 ACAAAACCTTGAAACCACTGGCTTACGCCCACTCTCAGAGGTTCCATGGGCGCAGG 1020  
961 ACAAAACCTTGAAACCACTGGCTTACGCCCACTCTCAGAGGTTCCATGGGCGCAGG 1020  
1021 GCCTCAGCCGTGCGTCTGCTGGGGCCAGGCCAGACCCACCTCGCAACATGTTTTC 1080  
1021 GCCTCAGCCGTGCGTCTGCTGGGGCCAGGCCAGACCCACCTCGCAACATGTTTTC 1080  
1081 TTGGCCCTGGGTAGTACATAGATGAGCAACTTTTAAACTTGGCATATTTCAATGTAATA 1140  
1081 TTGGCCCTGGGTAGTACATAGATGAGCAACTTTTAAACTTGGCATATTTCAATGTAATA 1140  
1141 GTCCAGATCCCGAGATCTTGTGAAGATGGCCATCCGGCCACAGCGGCTTCTATGGC 1200  
1141 GTCCAGATCCCGAGATCTTGTGAAGATGGCCATCCGGCCACAGCGGCTTCTATGGC 1200  
1201 TTGGTCTCTGGGATGTGCGCTTCTGTTCTGAGGGAACCCACCTCACCCGTCCTG 1260  
1201 TTGGTCTCTGGGATGTGCGCTTCTGTTCTGAGGGAACCCACCTCACCCGTCCTG 1260  
1261 CTTGCTGACCTGGAGCTGGAGCTGGCTCTCTCAGCTCTGCAAGTGTTCCTCCCTGC 1320

1261 CTTGCTGACCTGGAGGCTGGAGCTGGCTCTCTCCACCTCTGCAAGTGTTCCTCCCTGC 1320  
1321 AAATGCTCAAGGCTGTGTGGGCCAAGCCCGGATCGAAGCTGGAGCGTGAAGAAATGG 1380  
1321 AAATGCTCAAGGCTGTGTGGGCCAAGCCCGGATCGAAGCTGGAGCGTGAAGAAATGG 1380  
1381 GGAGGCTGGAGGCTGCCCCCAAGAGGCCACAGCCCTGGGAAGGGTCTGSCCTCTCTGGGG 1440  
1381 GGAGGCTGGAGGCTGCCCCCAAGAGGCCACAGCCCTGGGAAGGGTCTGSCCTCTCTGGGG 1440  
1441 CCAAGATGGGTGCCCCACCGTGCAGGAGTGGCCGAGGGTGGGATCGAGATCAGGAAG 1500  
1441 CCAAGATGGGTGCCCCACCGTGCAGGAGTGGCCGAGGGTGGGATCGAGATCAGGAAG 1500  
1501 GTTTTGGGCAAGAGCTAGCTGGAAGCTGAGCTCTGTGTGGAACCTCAGTTCAAAATCTTCC 1560  
1501 GTTTTGGGCAAGAGCTAGCTGGAAGCTGAGCTCTGTGTGGAACCTCAGTTCAAAATCTTCC 1560  
1561 TGTGTGAGGGCGGTGATGGTAGACTCAGCTCTGTGTGGAACCTCAGTTCAAAATCTTCC 1620  
1561 TGTGTGAGGGCGGTGATGGTAGACTCAGCTCTGTGTGGAACCTCAGTTCAAAATCTTCC 1620  
1621 CAGTGGCTGTAGAGTTGCTCTGACCACTAGAGGGCGCCCAACAGCATTAACCTGG 1680  
1621 CAGTGGCTGTAGAGTTGCTCTGACCACTAGAGGGCGCCCAACAGCATTAACCTGG 1680  
1681 GTCTGCCCTTCTTAGGACAAACCCACAGCTAGAGCTGTGCTGCTGCTGCTGCTGCTGCTG 1740  
1681 GTCTGCCCTTCTTAGGACAAACCCACAGCTAGAGCTGTGCTGCTGCTGCTGCTGCTGCTG 1740  
1741 CTTACTAGTTCTTTGGGTTCATGGAATTTACAAGCTTCTTAAGGAGCAGAGTGGCTCAG 1800  
1741 CTTACTAGTTCTTTGGGTTCATGGAATTTACAAGCTTCTTAAGGAGCAGAGTGGCTCAG 1800  
1801 ATTGGGGAAGCTGCGAGCTGTTCTCAGATCTGCAAAAGCGGTGTGTGGAGTATTGG 1860  
1801 ATTGGGGAAGCTGCGAGCTGTTCTCAGATCTGCAAAAGCGGTGTGTGGAGTATTGG 1860  
1861 TGAATCAAAAGAGAGGTTTGGCTGTAGTGGCCAGTCTTTAACTTAGATGCTCAGGGCC 1920  
1861 TGAATCAAAAGAGAGGTTTGGCTGTAGTGGCCAGTCTTTAACTTAGATGCTCAGGGCC 1920  
1921 GGGTGGGTATAAAATAAAGTAGGCTTTGAGCTGTGAGGCTTTGGGACTTTAAATTTT 1980  
1921 GGGTGGGTATAAAATAAAGTAGGCTTTGAGCTGTGAGGCTTTGGGACTTTAAATTTT 1980  
1981 TCCCACTATTCTCGAGATGGACATAGAGACATTTGCTTTGTGCTGAGAAATACTTGC 2040  
1981 TCCCACTATTCTCGAGATGGACATAGAGACATTTGCTTTGTGCTGAGAAATACTTGC 2040  
2041 ATGATGAGTCTGAGTCTGCTAAGGGCAACTGGCTTGAAGTGAACATCAAGGGGTGGTGGG 2100  
2041 ATGATGAGTCTGAGTCTGCTAAGGGCAACTGGCTTGAAGTGAACATCAAGGGGTGGTGGG 2100  
2101 ACTGTGGCAACCAACAGATTTCCACCTGAAATTTGGTGGCTGTCTTGGTGGGGCTAAT 2160  
2101 ACTGTGGCAACCAACAGATTTCCACCTGAAATTTGGTGGCTGTCTTGGTGGGGCTAAT 2160  
2161 GGGTGTACACCGAGATGTAGTAGGTCTGCTAAATGGGAGAGCTCTGGAGAACCCAGA 2220  
2161 GGGTGTACACCGAGATGTAGTAGGTCTGCTAAATGGGAGAGCTCTGGAGAACCCAGA 2220  
2221 AGCTAGATTTTTCATGTGAACCTATCCGAGTCTTAAAGTGTGTGAGCTAATGAGAAAA 2280  
2221 AGCTAGATTTTTCATGTGAACCTATCCGAGTCTTAAAGTGTGTGAGCTAATGAGAAAA 2280  
2281 CCTCTTAAACCTGTAGTCAAAAGGTGGGGCCATCTTTGACACCTCCCCACCAT 2340  
2281 CCTCTTAAACCTGTAGTCAAAAGGTGGGGCCATCTTTGACACCTCCCCACCAT 2340  
2341 GGTCTCTCAGGGACAGTGGCCCCCATGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 2400  
2341 GGTCTCTCAGGGACAGTGGCCCCCATGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 2400

QY 2401 GAACTGCTCAGGTGGCGGCTTGTGGCGCAGCACCCATCGTGGCTCCAGGTGTGGGCT 2460  
DB 2401 GAACTGCTCAGGTGGCGGCTTGTGGCGCAGCACCCATCGTGGCTCCAGGTGTGGGCT 2460  
QY 2461 GGGCTGGGCTATGGGTGGGCTCTACATGCTACAATAATGGGCTCATGATAAAAAAAA 2520  
DB 2461 GGGCTGGGCTATGGGTGGGCTCTACATGCTACAATAATGGGCTCATGATAAAAAAAA 2520  
QY 2521 AAAAAAAAAAAAAAAAAA 2538  
DB 2521 AAAAAAAAAAAAAAAAAA 2538

## RESULT 2

US-10-252-157-153  
; Sequence 153, Application US/10252157  
; Publication No. US20030190640A1  
; GENERAL INFORMATION:  
; APPLICANT: Pearson, Cecelia I.  
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER  
; FILE REFERENCE: PA-0027-1 US  
; CURRENT APPLICATION NUMBER: US/10/252,157  
; PRIOR FILING DATE: 2002-10-01  
; PRIOR APPLICATION NUMBER: 60/295,048  
; NUMBER OF SEQ ID NOS: 501  
; SOFTWARE: PERL Program  
; SEQ ID NO 153  
; LENGTH: 2461  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030190640A1 234056.5  
US-10-252-157-153

Query Match 41.8%; Score 1061.6; DB 12; Length 2461;

Best Local Similarity 72.8%; Pred. No. 3.3e-299;

Matches 1775; Conservative 0; Mismatches 524; Indels 139; Gaps 26;

QY 12 GCCGAGCGCGGCTAGGCGCGGCTCTAGCCAGCGCGCGCGCTGGAGGCGG 71  
DB 5 GCCGAGCGCGGCTAGGCGCGGCTCTAGCCAGCGCGCGCGCTGGAGGCGG 64  
QY 72 ATCCGG--CCCCGCTCGGTTCCCGGCGCGCGCGGCTGTCCATGCGGGCAA 129  
DB 65 GGCTGGCTCCGGCTCCGGTTCCCGGCGCGCGGCTGTCCATGCGCGCAA 124  
QY 130 GCACAGCACTTCCAGGAACCGAGTCCGCTGCTGCGGGAATACTTCTGTTGGCTT 189  
DB 125 GCACAGCATTTCCAGGAACCTGAGTCCGCTGCTGCGGGAATACTTCTGTTGGCTT 184  
QY 190 CAACATGTTTCTGGGTGGAGCCCTGTCTCTGCGCATCGGCTCTGGGCTGGG 249  
DB 185 CAACATGTTTCTGGGTGGAGCCCTGTCTCTGCGCATCGGCTCTGGGCTGGG 244  
QY 250 TGAGAGGCTGTTCTTCCACATCTCTGCGCTGACCGCATCTGGAGGCTTCGACCTGT 309  
DB 245 TGAGAGGCTGTTCTTCCACATCTCTGCGCTGACCGCATCTGGAGGCTTCGACCTGT 304  
QY 310 GTGGCTGTTTGTAGTGTGGAGCGCTCATGCTCGGCTGCTGGGCTTGTGGCTGCAITGG 369  
DB 305 GTGGCTGTTTGTAGTGTGGAGCGCTCATGCTCGGCTGCTGGGCTTGTGGCTGCAITGG 364  
QY 370 GGCTCTCGGAGAACACTTTCCTGCTCAAGTATTTCTCAGTGTTCCTTGGCTCATCTT 429  
DB 365 GGCTCTCGGAGAACACTTTCCTGCTCAAGTATTTCTCAGTGTTCCTTGGCTCATCTT 424  
QY 430 CTTCTCGAGCTGGCAACAGGATCTTGGCTTTCGATTTCAAGACTGGATTCGAGACCA 489  
DB 425 CTTCTCGAGCTGGCAACAGGATCTTGGCTTTCGATTTTCGATTTTCGAGACCA 484

QY 490 GCTCAATTTCTTCAATTAACAACAGTCAAGCGCTATCGGATGACATTTGACCTCCAGAA 549  
DB 485 GCTCAACCTTCTTCAATTAACAACAGTCAAGCGCTATCGGATGACATTTGACCTCCAGAA 544  
QY 550 CTTCAATTTGACTTTGCTCAGGAATAATTTGCTTCTGCGGAGCCGAGGCGCTAATGACTG 609  
DB 545 CTTCAATTTGACTTTGCTCAGGAATAATTTGCTTCTGCGGAGCCGAGGCGCTAATGACTG 604  
QY 610 GAACTCAATATCTATTTCAACTGCACTGTGAACCCGAGCGCAGAGCGCTGCGGGT 669  
DB 605 GAACTCAATATCTATTTCAACTGCACTGTGAACCCGAGCGCAGCGGCTGCGGGT 664  
QY 670 GCCCTTCTCTGCTGTGTCAGGAGCCCTGC-----GATGCTCTCAACACCGAGTGTGGCTA 725  
DB 665 GCCCTTCTCTGCTGTGTCAGGAGCCCTGCAGGAGATGCTCTCAACACCGAGTGTGGCTA 724  
QY 726 TGATGTCCGGCTCAAACTGGAGCTGGAGCAGGAGCTCCATACACACCAAGGCTGTGT 785  
DB 725 CGAGCTCCGGCTCAAACTGGAGCTGGAGCAGGAGCTTCATCCACACCAAGGCTGTGT 784  
QY 786 GGGCAGTTTGAGAAAGTGGCTGACGAGCAACCTGATCGTGGTGGCTGCGGTCTTTGTGGG 845  
DB 785 GGGCAGTTTGAGAAAGTGGCTGACGAGCAACCTGATTTGTTGGTGGGAGTCTTTCATGGG 844  
QY 846 CATGCTCTCTCCAGATCTTTGGTATCTGCTGCGCCAGAACTTGTGAGTGACATCAA 905  
DB 845 CATGCTCTCTCCAGATCTTTGGGATCTGCTGCGCCAGAACTTGTGAGTGACATCAA 904  
QY 906 GGCAGTGAAGGCCAACTGGATCAAACTGATGCTGCTACAACTACTCAAAATAACAAA 965  
DB 905 GGCAGTGAAGGCCAACTGGAGCAATGGAATGATG-----939  
QY 966 ACCTTGAACCACTGGCTTAGCGCCACCATCTCAGAGGTTCCATGGCGCGAGGCGCTC 1025  
DB 940 ACTTGAACCACTGGCTTAGCGCCACCATCTTCCAGAGTCTGCTCCACGCGGCGCTC 999  
QY 1026 AGCGTGGCTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1083  
DB 1000 AGCAGAACTCTGTGACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1059  
QY 1084 GCCTGGGTAGTACATACATGAGTGAAGCAACCTTTTAAACTTTGGCATATTTTCATGTAAGTC 1143  
DB 1060 GCCTGGGTAGTATACCTGAGCCCACTTTTAAATAATTTGAGATTTTACATAAAGTC 1119  
QY 1144 CAGATCCCGAGCATCTTGTGAAGATGGCCATCCGCGCCACAGCGGCTCTTCTATGGCTTC 1203  
DB 1120 CAGATCCACAGCTTCTTGAAGATGACACCTGCTGCTACGCGGCTCTTCCGTGGCAAC 1179  
QY 1204 GTCTCTGGGATGTGGCTTCTGTTCTGAGGAGCCACCTCACCGGTGTCC-----1258  
DB 1180 ACTACCTGGGACACTGCTCTCCAGTCAACGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1239  
QY 1259 ----TGCTGCTGCTGACCTGAGGCTGGGAGCTGGGCTCTCCCTCCCTCTGCAAGTTTTC 1314  
DB 1240 TAAGTGGCGCTGACCTTTGACTAGGAGCTGGGCTCT--CCACTCTGCAAGGTTAT 1297  
QY 1315 CCCTGCAAAATGCTGCAAGGCTGCTGTGGGCGCAAGCCCGGATCGAAGCTCGAGCGCTGAAG 1374  
DB 1298 TCCTGCA----CCTCGAGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1353  
QY 1375 AATTGGGAGGCTGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1434  
DB 1354 GA-TGGAGAGGCTGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1410  
QY 1435 TGGGCGGCGGAGAGTGGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1494  
DB 1411 TGGGAGCTGGGTGGGAG-TAGAGTGGCGGAGGAGGAGGCTGTGAGGCGGCGGCGGCGGCGG 1469  
QY 1495 AGGAAGGTTTGGGCGGAGGCTAGCTGGAAGCGCTGAGCTTGTCTACCCATGGGATGGGGA 1554  
DB 1470 AGGACAAATTTTGCAAAAGAGTAGCTGGAGGCG-----ATGGGACTGGCG 1514

1555 GAGCCCTGTTGAGGCGCTGATGTTAGGACTCAGCCCTCTGTTGGAATCAGTTCAAAA 1614  
Db GAGCCCTGTTGAGGCGCTGATGTTAGGACTCAGCCCTCTGTTGGAATCAGTTCAAAA 1566  
1615 TCTTCCAGTGGCCCTGTTAGGTTGCTCTGACCACTAGAGGGCGGCCACACAGCACTT 1674  
Db TCTTCCAGTGGCCCTGTTAGGTTGCTCTGACCACTAGAGGGCGGCCACACAGCACTT 1625  
1675 ACCTGGGTGCTCTTCTAGAGCAACCCACCCAGTACAGCCCTGTCCTGGTGTGTC 1734  
Db CCCTGGGTGCTCTTCTAGAGCAACCCACCCAGTACAGCCCTGTCCTGGTGTGTC 1672  
1735 ACCCTGCTTACTAGTCTTCTGGGTTTCATGGAATTTACAAGCTTCTAAGAGGAGAGTG 1794  
Db --CCTGCTACTAGCCTCTGGGTTTCGAGAGTTTGGGAATTTCTCA--GAGCCAATG 1728  
1795 GCTCAGATTGGGAAGCCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGT-----GT 1846  
Db GCTCAGATTGGGAAGCCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGT-----GT 1788  
1847 GTGCGAGTATTTGTGAATCAAGAGAGGTTTGGCTAGTCCCGAGTCTTTTAACCTAG 1906  
Db GTGCGAGTATTTGTGAATCAAGAGAGGTTTGGCTAGTCCCGAGTCTTTTAACCTAG 1848  
1907 ATGCCCTCAGGCGCGGTTGGTTTATAAAA-----ATAAGTAGGCTTTGAGCTGTGAGGC 1962  
Db ATGCCCTCAGGCGCGGTTGGTTTATAAAA-----ATAAGTAGGCTTTGAGCTGTGAGGC 1908  
1963 CTTTGGGACTTTAAATTTTCCCACTATTCCTGGAGATGGACATAGAGAGCAATGCTTT 2022  
Db TGTCTCACTTTGA--CTTTCCCACTGTGCTGGAG-----CAAGACATC 1952  
2023 GTGCTGAGAAATCTTGCATGATTCGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2082  
Db GTGCTGAGAAATCTTGCATGATTCGATGCTGATGCTGATGCTGATGCTGATGCTGATG 1953  
2083 CATCAAGGGGTGGTGGGACTGTGGCAACCAAGATCCCACTGAAATGGTGGCTGT 2142  
Db AAGAGAGTGGTGGGAGTGGGGGAGACAGGCTCCCGTCTGACAGTGGCGCTGC 2067  
2143 C--CTTCGTTGGGCTAATGCTGTACAGCAGAAATGATGATGATGATGATGATGATG 2200  
Db CATATTCAGGTGTAGCTAATTTCTGCTGGTGGGAATGCAG-----GCCTAATGACA 2119  
2201 GAACTCTGAGAGCAAGAGTCTAGATTTTTCATGCTGAACTATCCCGAGTTTAAAGTTG 2260  
Db GAAATCTGAGAGCAAGAGTCTAGATTTTTCATGCTGAACTATCCCGAGTTTAAAGTTG 2178  
2261 TTTGAGCTAATGAGAAACCTCTTAAACCTGTATAGTCAAAAGGTGTGGGGCCATCT 2320  
Db TTTGAGCTAATGAGAAACCTCTTAAACCTGTATAGTCAAAAGGTGTGGGGCCATCT 2179  
2321 TTGACACCTCCCAACCATAGTCTCCTCAGGACAGTCCCAAGGTGCAGGACCCAGCT 2380  
Db TTGACACCTCCCAACCATAGTCTCCTCAGGACAGTCCCAAGGTGCAGGACCCAGCT 2296  
2381 CCACGGTTTCAAGATGCTGGAACTGCTGCAGTGGGC 2418  
Db CTGCTATGAGGATGCTGGGCACTGCTGCCGGGTGGC 2334

## RESULT 3

US-09-934-268-1  
; Sequence 1, Application US/09934268  
; Patent No., US20020172986A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY  
; FILE REFERENCE: 10448-079001  
; CURRENT APPLICATION NUMBER: US/09/934,268  
; PRIOR FILING DATE: 2001-08-21  
; APPLICATION NUMBER: 60/226,612

PRIOR FILING DATE: 2000-08-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3184  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (168)...(977)  
US-09-934-268-1

Query Match 31.1%; Score 790.4; DB 10; Length 3184;  
Best Local Similarity 92.3%; Pred. No. 7.2e-220;  
Matches 855; Conservative 0; Mismatches 66; Indels 5; Gaps 2;

QY 4 CGCGTCCCGCGCAGCCCGCGGCTAGGCCCGCGGCTCTAGCCCGAGGGCGGCCGCTG 63  
Db CGCATGAAGCCGAGCCCGCGGCTAGGCCCGCGGCTCTAGCCCGAGGGCGGCCGCTG 111  
QY 64 GAGGCGCATCCCGGC-CCCGGCTCCGTTCCCGGCGCGGCGGCTGTCTCACCATGC 122  
Db GGGCGCTGGGCTGGCTCCCGCTCCGCTCCCGGCGCGGCGGCTGTCTCACCATGC 171  
QY 123 CGGCAAGCAGCAGCACTTCCAGGAACCGAGGTCGGTCTGCGGGAATACTTCTCTGT 182  
Db CGGCAAGCAGCAGCACTTCCAGGAACCGAGGTCGGTCTGCGGGAATACTTCTCTGT 231  
QY 183 TTGGCTTCAACATTTCTGGGTCGCGAGCCCTGTTCTGCGCCATCGGCTCTGGG 242  
Db TTGGCTTCAACATTTCTGGGTCGCGAGCCCTGTTCTGCGCCATCGGCTCTGGG 291  
QY 243 CCGTGGGTGAGAAAGGTGTTCTCTCAACATCTCTGCGCTGACCGATCTGGAGGCTCG 302  
Db CCGTGGGTGAGAAAGGTGTTCTCTCAACATCTCTGCGCTGACCGATCTGGAGGCTCG 351  
QY 303 ACCCTGTGGCTGTTTGTAGTGGTGGAGGCTCATGTCCGTGCTGGGCTTTGCCGCT 362  
Db ACCCGTGTGGCTGTTTGTAGTGGTGGAGGCTCATGTCCGTGCTGGGCTTTGCCGCT 411  
QY 363 GCATCGGGCTCTCGGGAGAACACTTCTCTGCTCAAGTCTTCTCAGTGTCTCTGGCC 422  
Db GCATCGGGCTCTCGGGAGAACACTTCTCTGCTCAAGTCTTCTCAGTGTCTCTGGCC 471  
QY 423 TCATCTTCTTCTGAGCTGGCAACAGAGATCTTGGCTCTTGTATTTCAAGGACTGGATTC 482  
Db TCATCTTCTTCTGAGCTGGCAACAGAGATCTTGGCTCTTGTATTTCAAGGACTGGATTC 531  
QY 483 GAGACAGCTCAATTTCTTCAATTAACACACGTCAGGCTTATCGGATGACATTCACC 542  
Db GAGACAGCTCAATTTCTTCAATTAACACACGTCAGGCTTATCGGATGACATTCACC 591  
QY 543 TCCAGAACCTCAATTTCTGCTCAGGAATATTGGTCTTGTGCGGAGCCCGAGGCTCTA 602  
Db TCCAGAACCTCAATTTCTGCTCAGGAATATTGGTCTTGTGCGGAGCCCGAGGCTCTA 651  
QY 603 ATGACTGGAACCTCAATTTCTTCAACTGCACTGATTTGAACCCGAGCCGAGAGCGCT 662  
Db ATGACTGGAACCTCAATTTCTTCAACTGCACTGATTTGAACCCGAGCCGAGAGCGCT 711  
QY 663 GCGGGTGGCTTCTCTGCTGTGTGTCAGGAGCCCTGCG-----GATGTCTCAACCCAGT 718  
Db GCGGGTGGCTTCTCTGCTGTGTGTCAGGAGCCCTGCGGGAGGATGTCTCTCAACCCAGT 771  
QY 719 GTGGCTATGATGTCGGCTCAAACTGGAGCTGGAGCAGCAGGCTTCCATACACCAAG 778  
Db GTGGCTATGATGTCGGCTCAAACTGGAGCTGGAGCAGCAGGCTTCCATACACCAAG 831  
QY 779 GCTGTGGGCAAGTTTGAAGTGGTGTGAGGACAACTGATCTGTGTGGTGGGCTCT 838  
Db GCTGTGGGCAAGTTTGAAGTGGTGTGAGGACAACTGATCTGTGTGGTGGGCTCT 891  
QY 839 TTGTGGGCACTGCTCTCTCCAGATCTTTGGTATCTGCTGCGCCAGAACCTTGTGAGTG 898

[illegible]

## RESULT 4

```

US-10-162-435-35
; Sequence 35, Application US/10162435
; Publication No. US20030096305A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Curtis , Rory A. J.
; APPLICANT: Kapeller-libermann, Rosana
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Leiby , Kevin R.
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND
; TITLE OF INVENTION: CELL SURFACE PROTEIN FAMILY MEMBERS
; FILE REFERENCE: 10448-189001
; CURRENT APPLICATION NUMBER: US/10/162,435
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/836,499
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: PCT/US01/12420
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 60/197,507
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/891,008
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: PCT/US01/19963
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/214,220
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/860,868
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/16013
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,674
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/886,429
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: PCT/US01/20055
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/213,963
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 10/041,406
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: PCT/US02/00275
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/260,286
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: US 09/934,268
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/41811
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,612
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 3184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (168) ... (977)
US-10-162-435-35

```

RESULT 5  
US-10-103-196-4  
; Sequence 4, Application US/10103196  
; Publication No. US20030050466A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.



```
; TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT008PI
; CURRENT APPLICATION NUMBER: US/10/103,196
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/707,936
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/13504
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/178,770
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/149,447
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/138,573
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/137,797
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/135,122
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1126
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-196-4

Query Match      31.0%; Score 787.2; DB 14; Length 1126;
Best Local Similarity 92.6%; Pred. No. 3.6e-219;
Matches 850; Conservative 0; Mismatches 63; Indels 5; Gaps 2;

Qy 12 GCCCGAGCCGCCGCTAGGCGCCGCGGCTAGCCAGCGGCGGCGGCGGCGGCGGCGG 71
Db 49 GCCCGAGCCGCCGCTAGGCGCCGCGGCTAGCCAGCGGCGGCGGCGGCGGCGGCGG 108
Qy 72 ATCCCGGC -CCCGGCTCCGGTTCGCCGGCCGCGGCGGCGGCTGCTCACCATGCCCGGCAAG 130
Db 109 GGCCTGGCTCCCGGCTCCGGTTCGCCGGCCGCGGCGGCTGCGGCTCACCATGCCCGGCAAG 168
Qy 131 CACGAGCACTTCCAGGAACCCGAGTCCGGCTGCTGCGGGAATACTTCTCTTTGGGTTTC 190
Db 169 CACGAGCACTTCCAGGAACCCGAGTCCGGCTGCTGCGGGAATACTTCTCTTTGGGTTTC 228
Qy 191 AACATTGTTTCTGGGTCTGGAGCCCTGTTCTTGGCCATCGGCGCTGCGGCGTGGGT 250
Db 229 AACATTGTTTCTGGGTCTGGAGCCCTGTTCTTGGCTATCGGCGCTTGGGCGTGGGT 288
Qy 251 GAGAGGGGTGTTCTTCAACATCTCTGCGTACCGATCTGGAGGCGCTCGACCCCTGTG 310
Db 289 GAGAGGGCGTCTCTCGAACATCTCAGCGCTGACAGATCTGGAGGCGCTTGACCCCGTG 348
Qy 311 TGGCTGTTTGTAGTGGTGGAGGGGTGATGTCGCTGCTGGCGCTTGGCGGCTGATCGGG 370
Db 349 TGGCTGTTTGTAGTGGAGGGGTGATGTCGCTGCTGGGCTTTGTGCTGCTGATTTGGG 408
Qy 371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTCTCAGTGTTCCTTGGCGCTCATCTTC 430
Db 409 GCGCTCCGGGAGAACACTTTCCTGCTCAAGTTTTCTCAGTGTTCCTTGGCGCTCATCTTC 468
Qy 431 TTCTCGAGCTGGCAACAGGGATCTTGGCCCTTCGTATTCAAGGACTGGATTCGAGACCAAG 490
Db 469 TTCTCGAGCTGGCAACAGGGATCTTGGCCCTTCGTATTCAAGGACTGGATTCGAGACCAAG 528
Qy 491 CTCGAATTTCTTATTAAACAAAGTCAAGGCTTATCGGATGACATTTGACCTCCAGAAC 550
Db 529 CTCGAATTTCTTATTAAACAAAGTCAAGGCTTATCGGATGACATTTGACCTCCAGAAC 588
Qy 551 CTCATTGACTTTTGTCTAGGAATAATTTGTTCTTGTCTGGAGCGCGGCGGCTTAATGACTGG 610
Db 589 CTCATTGACTTTTGTCTAGGAATAATTTGTTCTTGTCTGGAGCGCGGCGGCTTAATGACTGG 648
Qy 611 AACCTCAATATCTATTATTTCACTGCACTGACTTGAACCCGAGCGGAGCGCTCGGGGGTG 670
Db 649 AACCTCAATATCTATTATTTCACTGCACTGACTTGAACCCGAGCGGAGCGCTCGGGGGTG 708
```

```
Qy 671 CCCTTCTCTGCTGTGTCAAGGACCCTGCG-----GATGTCCTCAACACCCAGTGTGGCTAT 726
Db 709 CCCTTCTCTGCTGTGTCAAGGACCCTTGGCGAGGATGTCTCAACACCCAGTGTGGCTAC 768
Qy 727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACAAAGGCTGTGTG 786
Db 769 GACGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCATCCACACAAAGGCTGCGTG 828
Qy 787 GCCCAGTTTGAAGAGTGGCTCAGGACAACTGATCGTGGTGGCTGGGGTCTTTCTGGGC 846
Db 829 GGCAGTTTGAAGATGGCTCAGGACAACTGATGTTGGTGGCGGAGTCTTTCATGGGC 888
Qy 847 ATCCTCTCTCTCCAGATCTTTTGGTATCTGCTGCCCGCAGAACTTGTGAGTGACATCAAG 906
Db 889 ATCGCCCTCTCTCCAGATCTTTTGGCATCTGCTGCCCGCAGAACTTGTGAGTGACATCAAG 948
Qy 907 GCAGTGAAGGCCAACTGG 924
Db 949 GCAGTGAAGGCCAACTGG 966

RESULT 6
US-09-934-268-3
; Sequence 3, Application US/09934268
; Patent No. US20020172986A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; TITLE OF INVENTION: A NOVEL HUMAN TETRASPANIN FAMILY
; TITLE OF INVENTION: MEMBER AND USES THEREOF
; FILE REFERENCE: 10448-079001
; CURRENT APPLICATION NUMBER: US/09/934,268
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/226,612
; PRIOR FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 813
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-934-268-3

Query Match      28.0%; Score 710.4; DB 10; Length 813;
Best Local Similarity 93.2%; Pred. No. 8.7e-197;
Matches 755; Conservative 0; Mismatches 51; Indels 4; Gaps 1;
```

```
Qy 119 ATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGTCCGCTGCTGCGGAAATACTTC 178
Db 1 ATGCCGGGCAAGCACCAGCACTTCCAGGAACCCGAGTCCGCTGCTGCGGAAATACTTC 60
Qy 179 CTGTTTGGCTTCAACATTTGTTTCTGGGTGCTGGAGCCCTGTTCTTGGCTATCGGCTC 238
Db 61 CTGTTTGGCTTCAACATTTGTTTCTGGGTGCTGGAGCCCTGTTCTTGGCTATCGGCTC 120
Qy 239 TGGGCTTGGGTGAGAGGGTGTCTCTCAACATCTCTGCGCTGACCGATCTCGGAGGC 298
Db 121 TGGGCTTGGGTGAGAGGGTGTCTCTCAACATCTCTGCGCTGACCGATCTCGGAGGC 180
Qy 299 CTGACCCCTGTGTGGCTGTTTGTAGTGGTGGAGGGCTCATGCTCGTGTGGGCTTTGCC 358
Db 181 CTGACCCCTGTGTGGCTGTTTGTAGTGGTGGAGGGCTCATGCTCGTGTGGGCTTTGCC 240
Qy 359 GGCTGCAATCGGGGCTCTCCGGGAGAACACTTCTCTGCTCAAGTTTTTCTCAGTGTTCCTT 418
Db 241 GGCTGCAATGGGGGCTCTCCGGGAGAACACTTCTCTGCTCAAGTTTTTCTCAGTGTTCCT 300
Qy 419 GGCTCATCTTCTTCTCGAGCTGGCAACAGGATCTTGGCTTCTGTAATCAAGGACTGG 478
Db 301 GGCTCATCTTCTTCTCGAGCTGGCAACAGGATCTTGGCTTCTGTAATCAAGGACTGG 360
Qy 479 ATTGGAGACCACTCAATTTCTTCAATTAACAAACGTCGAGGCTTATCGGAGTGCATT 538
Db 1 ATTGGAGACCACTCAATTTCTTCAATTAACAAACGTCGAGGCTTATCGGAGTGCATT 538
```





Qy	135	AGCACTTCCAGGAACCCGAGGTGGCTGCTCGGGGAATATCTTCTGTTTGGCTTCAACA	194
Db	287	AGCACTACAAGGGTCTGAAGTCAGTTGTGTGCATCAATCTTCATATTTGGCTTCAATG	346
Qy	195	TTGTTTCTGGGTGCTGGGACCCGTGTCCTGGCCATCGGCGCTCTGGGCGTGGGTGAGA	254
Db	347	TCATATTTGGTTTTTGGGAATACATTTCTTGGNATGGACTGTGGCATGGAAATGAA	406
Qy	255	AGGTGTTCTCTCCAAACATCTCTGGCGTGAACGATCTGGGAGGCTCGACCCTGTGTGC	314
Db	407	AAGGAGTTCTGTCCAACATCTCTTCATCACCGATCTCGGGGCTTTGACCCAGTTTGGC	466
Qy	315	TGTTTGTAGTGGTGGAGGCGTCATGTCGCTGGGCTTTGCGGCTGGATCGGGGCTC	374
Db	467	TCTTCTTGTGTTGGAGGAGTGATGTTCATTTTGGGATTTTGCAGGTGTCATTTGAGCGC	526
Qy	375	TCCGGGAGAACATTTTCTGTCTCAAGTTTCTCAGTGTTCCTTTGGCCTCATCTTCTTCC	434
Db	527	TACGGGAAAACATTTCTTCTCAAGTTTCTTCTGTGTTCTCTGGAAATATTTTCTTCC	586
Qy	435	TGGAGCTGGCAACAGGAGATCTTGCCCTTCGTATTTCAAGGACTGGATTCGAGACCACTCA	494
Db	587	TGGAGCTCACTCGCGGAGTCTTAGCATTTGTTTTCAAAGACTGGATCAAAGACCACTGT	646
Qy	495	ATTTCTTCAATTAACAACAACCTCAAGGCTTATCGGGATGACATTTCACTCCAGAACTCA	554
Db	647	ATTTCTTTATAAACAAACAACATCAGAGCATATCGGATGACATTTGTAATTTGCAAACTCA	706
Qy	555	TTGACTTTGCTCAGGAATATTTGGTCTTGCTCGGAGCCGAGGGCTTAATGACTTGBAAC	614
Db	707	TAGACTTCACCCAGGAATATTGGCAGTGCTGTGGGGCTTTTGGAGCTGATGATTTGSAAC	766
Qy	615	TCAATATCTATTTCAACTGCACTTGAAACCGAGCCGAGAGCGCTCGGGGTGCCT	674
Db	767	TAAATATTTACTTCAATTTGCACAGATCCCAATGCAAGTCGAGAGGATGTGGCGTTCCAT	826
Qy	675	TCTCTGCTGTGCAGGACCCCTGC - - - GATGTCTCAACACCCAGTGTGGCTATGATG	730
Db	827	TCTCTGCTGCATAAAGATCCCGAGAAGATGTCATCAACACTCAGTGTGGCTATGATG	886
Qy	731	TCCGGCTCAAACTGGAGCTGGAGCAGAGGCTCCATACACCAAAAGGCTGTCTGGGCC	790
Db	887	CCAGGCAAAACAGAGATTGACCAGCAGATTGTAACTACACGAAGGCTGTGTGCCCC	946
Qy	791	AGTTTGAGAAGTGTGCTCAGGACAACTGATCTGTGGTGGCTTCTTTGTGGGCACTG	850
Db	947	AGTTTGAGAAGTGTGTCAGGACAAATTAACCATCGTGTCTGGTATTTTTCATAGGCATG	1006
Qy	851	CTCTCTCCAGATCTTTTGGTATCTGCTGGCCCAAGAACCTTGTGAGTGACATCAAGCAG	910
Db	1007	CATTGCTGCAGATATTTGGGATATGCTTGGCCCAAGATTTTGGTTAGCGATATCGAAGCTG	1066
Qy	911	TGAAGGCCAACTGG	924
Db	1067	TCAGGGCGAGCTGG	1080

RESULT 10  
US-10-103-196-5  
; Sequence 5, Application US/10103196  
; Publication No. US20030050466A1  
; GENERAL INFORMATION:  
; APPLICANT: NI et al.  
; TITLE OF INVENTION: TMS4F Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PT008P1  
; CURRENT APPLICATION NUMBER: US/10/103,196  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/707,936  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: PCT/US00/13504  
; PRIOR FILING DATE: 2000-05-18  
; PRIOR APPLICATION NUMBER: 60/178,770

```

; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/149,447
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/138,573
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/137,797
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/135,122
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-103-196-5

```

Query Match	16.5%	Score 418.6	DB 14	Length 1174
Best Local Similarity	73.0%	Pred. No. 2.1e-111		
Matches 580	Conservative 0	Mismatches 209	Indels 6	Gaps 3
Qy	135	AGCACTCCAGGAACCCGAGCTCGGCTGCTGCGGGAATACATTCCTGTTTGGGCTTCAACA	194	
Db	103			
Qy	195	TTGTTTTCTGGGTGCTGGGAGCCCTGTCCTCGGCCATCGGCCTCGGGCTGGGGTGAGA	254	
Db	163	TCATATTTTGGTTTTGGGAATACATTTCTTGGAAITGGACTGTGGGCATGAAGTGA	222	
Qy	255	AGGGTGTCTCTCCAAATCTCTCGCGCTGACCGATCTGGAGGCTCGACCTCTGTGTGGC	314	
Db	223	AAGGAGTCTGTCCAAATCTCTTCCATCACCGATCTCGCGGCTTTGACCCAGTTTGGC	282	
Qy	315	TGTTTGTAGTGTGGAGGGCTCATGTCGTGCTGGGCTTTGCGGCTGCATCGGGGCTC	374	
Db	283	TC TTCTCTGTGTGGGAGGAGTGATGTTCAATTTTGGGATTTGCAGGGTGCAITGGAGC-G	341	
Qy	375	TCGGGAGAACATTTCTCGCTCAAGCTTTTCTCAGTGTTCTCTGGGCTCATCTTC-TTC	433	
Db	342	TACGGGAAAACACTTCTCTCAAGTTTTTTTCTGTGTCTGGGAAATATTTTCTTTTC	401	
Qy	434	CTGGAGCTGGCAACAGGGATCTTTGGGCTTCGTATTCAAGGACTGGGATTCGAGACCAAGCTC	493	
Db	402	CTGGAGCTCACTGCCGAGTTCTAGCATTTGTTTTCAGAGCTGGATCAAGACCAAGCTG	461	
Qy	494	AATTTCTTCAATTAAACAAACAGTCAAGGCTATCGGGATGACATTGACCTCCAGAACCTC	553	
Db	462	TATTTCTTTATAAACAAACACATCAGAGCATATCGGGATGACATTGATTTGCAAAACCTC	521	
Qy	554	ATTGACTTTGCTCAGGAATATTTGTTCTGCTGGAGCCCGAGGGCTAATGACTGGAAAC	613	
Db	522	ATAGACTTCACCCAGGAATATTTGGCAGTGTCTGCGGGCTTTTTTGGAGCTGATGATTTGGAAAC	581	
Qy	614	CTCAATATCTATTTCAACTGCACTGCACTTGAACCCGAGCCGAGAGCGCTCGCGGGTGCCC	673	
Db	582	CTAATATTTACTTTCAATTGCAAGATTTCCATGCAAGTTCGAGAGCGATGTGGGTTCCA	641	
Qy	674	TTTCTCTGCTGTGTCAAGGACCCCTGTC- - -GATGTCTCTCAACCCAGCTGTGGCTATGAT	729	
Db	642	TTCTCTGCTGCTAAAGATCCCGAGAAGATGTCAATCAACACTCAGTGTGGCTATGAT	701	
Qy	730	GTCCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAGGCTGTGTGGGC	789	
Db	702	GCCAGGCAAAAACGAAGTTGACACAGAGATTTGTAATCTACACGAAGGCTGTGTGGCC	761	

QY 790 CAGTTTGGAGAGTGGCTGCAGGACAACTGATCGTGGTGGCTGGGTCTTTGTGGGCATC 849  
DB 762 CAGTTTGGAGAGTGGCTGCAGGACAACTTAACTCATCGTGGTGGTATTTTCATAGGCATT 821  
QY 850 GCTCTCTCCAGATCTTTGGTATCTGCTGGCCAGAACCTTTGTGAGTGACATCAAGGCA 909  
DB 822 GCATTGCTGCAGATATTTGGGATATGCTGCTGGCCAGAAATTTGGTTAGCGATATCAAGCT 881  
QY 910 CTGAGGCGCAACTGG 924  
DB 882 GTGAGGCGAGCTGG 896

RESULT 11  
US-10-103-196-10  
; Sequence 10, Application US/10103196  
; Publication No. US20030050466A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: TMASF Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PT008P1  
; CURRENT APPLICATION NUMBER: US/10/103,196  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/707,936  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: PCT/US00/13504  
; PRIOR FILING DATE: 2000-05-18  
; PRIOR APPLICATION NUMBER: 60/178,770  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/149,447  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/138,573  
; PRIOR FILING DATE: 1999-06-11  
; PRIOR APPLICATION NUMBER: 60/137,797  
; PRIOR FILING DATE: 1999-06-03  
; PRIOR APPLICATION NUMBER: 60/135,122  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 1178  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-103-196-10

Query Match 15.8%; Score 402.2; DB 14; Length 1178;  
Best Local Similarity 71.6%; Pred. No. 1.3e-106; Mismatches 218; Indels 8; Gaps 3;  
Matches 571; Conservative 0;

QY 135 AGCACTTCCAGGAACCCGAGGTGGCTGCTGGGAAATATCTCTCTGTTGGCTTCAACA 194  
DB 101 AGCACTCAAGGGTCTCTGAAGTCAGTTGTTGCATCAATATCTCATATTTGGCTTCAATG 160  
QY 195 TTGTTTCTGGTGGTGGAGCCCTGTTCTGCGCCATCGGCTCTGGGCTCGGGGTGAGA 254  
DB 161 TCATATTTGGTGGTGGGAAATCAATCTTTGGAATTTGCACTGTGGGCTGGAATGAA 220  
QY 255 AGGGTGTCTCTCCAACTCTCTGGCTGACCGATCTGGAGGCTCGACCTGTGTGC 314  
DB 221 AAGGAGTCTGTCCAACTCTCTCCATCACCAGATCTCGGGGCTTTGACCCAGTTTGGC 280  
QY 315 TGTGTTAGTGGTGGAGCGTCAATGCTGCGGCTTTGCGGCTGCACTCGGGGCTC 374  
DB 281 TCTTCCTTGTGGTGGAGGAGTGATGTTTCATTTGGGATTTGCAAGGTGCAATGGAGC-G 339  
QY 375 TCCGGAGAACACTTCTCTGCTCAAGTTTTTCTC---AGTGTCTTGGGCTCATCTTCT 431  
DB 340 TACGGGAAAACTTCTCTCTCAAGTTTTTCTCGGTGTTCTCGGGAATTAATTTCTT 399  
QY 432 TCCTGGAGCTGGCAACAGGAGTCTTGGGCTTCTGCTTATCAAGGACTGGATTTCGAGCAGC 491  
DB 400 TCCTGGAGCTCACTGCCGAGGTCTTAGCATTTGTTTCAAAGACTGGATCAAGACCCAGC 459

QY 492 TCAATTTCTTCATTAAACAACACGTCAGGCTTATCGGATGACATTTGACCTCCAGAAC 551  
DB 460 TGTATTTCTTTATTAACAACAACATCATGAGCATATCGGATGACATTTGATTTGCAAAAC 519  
QY 552 TCATTGACTTTGGCTCAGGAATATTGGTCTTCTGCTGGAGCCCGAGGCTTAATGACTGA 611  
DB 520 TCATAGACTTCAACCAGGAATATTGGCAGTCTGTGGGGCTTTTGGAGCTGATGATTGA 579  
QY 612 ACCTCAATATCTATTTTCACTGACTGACTTGAACCCGAGCCGAGAGCGCTCGGGGTGC 671  
DB 580 ACCTAAATATTACTTTCAATTTGCAAGATTCGAATTCGAAGTCGAGAGCGATGTGGCGTTC 639  
QY 672 CTTCTCTCTGCTGTGTGAGGACCTCTGC---GATGTCTCAACACCCAGTGTGGCTATG 727  
DB 640 CATCTCTCTGCTGCATTAAGATCCGAGAGAGATGTCATCAACTCAGTGTGGCTATG 699  
QY 728 ATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGCTCCATACACACCAAGGCTGTGTGG 787  
DB 700 ATGCCAGCAAAACCAAGTTGACCAAGTTGTAATCTACACGAAAGGCTGTGTGC 759  
QY 788 GCCAGTTTGAAGTGGCTGCAGGACCACTGATGCTGTGGTGGGCTCTTTTGGGCA 847  
DB 760 CCCAGTTTGAAGTGGTTCAGGACAAATTTAACCATCGTTGCTGGTATTTTCATAGGCA 819  
QY 848 TCGCTCTCTCCAGATCTTTGGTATCTGCTGGCCAGAACCTTGTGAGTGACATCAAGG 907  
DB 820 TTGCATCTCTGCAGATATTTGGGATATGCTTGGCCAGAAATTTGGTAGCGATATCGAAG 879  
QY 908 CAGTGAAGGCCAACTGG 924  
DB 880 CTGTGAGGCGAGCTGG 896

RESULT 12  
US-09-822-846-126  
; Sequence 126, Application US/09822846  
; Publication No. US20030027139A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LaVallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Bowman, Michael R.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakara  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6400  
; CURRENT APPLICATION NUMBER: US/09/822,846  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195,605  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 629  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 126  
; LENGTH: 1988  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-846-126

Query Match 15.4%; Score 390.8; DB 11; Length 1988;  
Best Local Similarity 74.3%; Pred. No. 3.7e-103;

Matches 508; Conservative 0; Mismatches 172; Indels 4; Gaps 1;	
QY	245 TCGGGTGAGAAGGCTGCTCTCCACATCTCTGCGCTGACCGATCGGAGCCCTCGAC 304
Db	2 TGAATGAANAAGAGTCTGTGCAACATCTCTCCATCACCAGATCTCGGCGCTTTGAC 61
QY	305 CCTGTGTGGCTGTTTGTAGTGTGTGAGCGTGCATGTCGTGTGCGCTTTGCGGCTGC 364
Db	62 CCAAGTTGGCTCTCTCTGTTGGTGGAGGATGATGTTCAATTTGGGATTTGAGGGTGC 121
QY	365 ATCGGGCTCTCGGAGAACACTTTCTGCTCAAGTTTCTCAGTGTTCCTTGGCCCTC 424
Db	122 ATTGGAGCGCTACGGGAACAACCTTTCTCTCAAGTTTCTCTGTTCTCTCGGAATT 181
QY	425 ATCTTCTTCTCGAGCTGCAACAGGGATCTTGGCTCTCGTATTCAGGACTGGATTCCA 484
Db	182 ATTTTCTTCTGAGCTACTCGCGAGTCTTAGCAATTTGTTTCAAGACTGGATCAAA 241
QY	485 GACCACTCAATTTCTTCAATTAACAACAGCGCTATCGGGATGACATTTGACCTC 544
Db	242 GACCACTGATTTCTTTATAACAACAACATCAGAGCATATCGGGATGACATTTGATTG 301
QY	545 CAGAACCTCATGACTTTCTCAGGAATATTTGCTGTGCGAGCCCGAGGCGCTAAT 604
Db	302 CAANAACCTCATGACTTCAACCAGGAATATTTGCGAGTGTGTTGGAGCTGAT 361
QY	605 GACTGGAACCTCAATCTATTTTCAACTGACTTGAACCCGAGCCGAGAGCGCTGC 664
Db	362 GATTGGAACCTAATATTTACTTCAATTCACAGATTCATGCAAGTCTGAGAGCGATG 421
QY	665 GGGGTGCCCTTCTCTGCTGTGTGAGGACCCCTGC ---GATGTCTCTCAACACCCAGTGT 720
Db	422 GCGGTTCATTTCTCTGCTGCACTAAAGATCCCGCAGAGATGTCAATCAACTCAGTGT 481
QY	721 GCTATGATGTCGGCTCAACTGGAGCTGGAGCAGGCTCCATACACACCAAGGC 780
Db	482 GCGTATGATGCCAGGCAAAACCAAGAGTTGACAGAGATTTGTAATCTACACGAAGGC 541
QY	781 TGTGTGGGCCAGTTTGAGAAGTGGCTGACAGGACCACTGATCGTGTGCTGGGCTTTT 840
Db	542 TGTGTGCCCACTTTGAGNAGTGTGTGAGGACAAATTTAAACATCGTGTCTGATTTTC 601
QY	841 GTGGCATGCTCTCTCCAGATCTTTGGTATCTGCTGGCCAGAACCTTGTGAGTGAC 900
Db	602 ATAGGCAITGCAATGCTGCAGATATTTGGGATATGCTGGCCAGAAATTTGTGTAGCAT 661
QY	901 ATCAGGCGAGTCAAGCCCACTGG 924
Db	662 ATCGAAGCTGTGAGGGCGAGCTGG 685

RESULT 13  
US-09-892-877-20  
; Sequence 20, Application US/09892877  
; Publication No. US20030077809A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et. al.  
; TITLE OF INVENTION: 97 Human secreted proteins  
; FILE REFERENCE: P2028pi  
; CURRENT APPLICATION NUMBER: US/09/892,877  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10  
; NUMBER OF SEQ ID NOS: 461  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 2672  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (16)  
; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE	
; LOCATION: (28)	
; OTHER INFORMATION: n equals a,t,g, or c	
; NAME/KEY: SITE	
; LOCATION: (47)	
; OTHER INFORMATION: n equals a,t,g, or c	
; NAME/KEY: SITE	
; LOCATION: (52)	
; OTHER INFORMATION: n equals a,t,g, or c	
; NAME/KEY: SITE	
; LOCATION: (93)	
; OTHER INFORMATION: n equals a,t,g, or c	
US-09-892-877-20	
Query Match	
Best Local Similarity 14.1%; Score 357.4; DB 11; Length 2672;	
Matches 564; Conservative 3; Mismatches 304; Indels 6; Gaps 2;	
QY	48 CCAGGGCGGCGCGTGGAGGGCGGATCCGGGCGCGGCTCCGGTCCCGGGCGGGCGGC 107
Db	99 CCGGGGCTGCGAGGAATTTTCAGCGGGGCGCGCCCGCTGCCCGCGCGGCGGC 158
QY	108 GGCTGCT--CACCATGCCGGCAAGCACGACATTTCCAGGAACCCGAGTCCGGCTGCTG 165
Db	159 GGATTTGCTTCTCAGAGATGCATATTTATAGATACTCTAAGCCCAAGTCAAGTCTG 218
QY	166 CGGGAATATCTTCTGTTGGTTCGAATGTTTCTGGGTGCTGGAGCCCTGTTCTTCT 225
Db	219 GTACAAGTACCTCTCTTTTTCAGCTACAAATCATCTTCTGTTGGCTGGAGTTGCTTCT 278
QY	226 GGCATCGGCTCTGCGGCTGGGTGAGAGGCTGTTCTCTCAACATCTCTCGGCTGAC 285
Db	279 TGAAGTGGGCTGTGGCATGAGGCAAGAGGTGCTGTCGACCTCACCAGAGTGC 338
QY	286 CGATCTGGAGGCTCGACCTCTGTGTGCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGT 345
Db	339 CGGATGCATGCAATCGACCTCTGTGTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 398
QY	346 GCTGGCTTTTGGCGCTGATCGGGCTCTCCGGAGAAACATTTTCTGCTCAAGTTTCT 405
Db	399 CCGGGGTTGCGCGCTGCTGCGGGGCTCTGCGGGAGAAATATCTGTGCTCACTTTT 458
QY	406 CTCAGTGTTCCTTGGCTCATCTTCTCTGAGCTGGCAACAGGATCTTGGCTTCTG 465
Db	459 CTGTGGCACTATGCTGCTCATCTTCTCTGAGCTGGCTGTGGCGCTGCTGGCTTCT 518
QY	466 ATTCAGGACTGGATTTCGAGACCACTCAATTTCTTCAATTAACAACAGCTCAAGCCTA 525
Db	519 GTTCAGGACTGGGTGAGGACCGGTTCCGGAGTTCTTCGAGAGCAACATCAAGTCTTA 578
QY	526 TCGGATGACATTTGACCTCCAGAACCTCATTTGCTTGTCTCAGGAATATTTGCTTGTG 585
Db	579 CCGGACGATATCGATCTGCAAAACCTCATCGACTCCCTTCAGAAAGCTAAACAGTCTG 638
QY	586 CGAGCCCGAGGCGCTTAATGACTGGAACCTCAATATCTATTTCAACTGCACTGACTGAA 645
Db	639 TGGCGCATATGCGCTTGAAGACTTGGACCTCAACGCTACTTCAATTTGAGCGGTGCCAG 698
QY	646 CCCAGCCGAGAGCGCTCGGGGCTGCTTCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 701
Db	699 CTACAGCCGAGAGAGTGTGGGGTCCCTTCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 758
QY	702 TGTCTTCAACACCCAGT 761
Db	759 AGTTGTGAACACACAGT 818
QY	762 CTCATACACACCAAGGCTGTGTGGGCACTTTTTCAGAGTGGCTGTGAGGACCACTGAT 821
Db	819 GTCCATCTTTCAGAAAGGCTGCATCCAGGCGCTGGAAGCTGGCTCCCGCGGAAACATTTA 878
QY	822 CGTGTGTGGCTGGGTCTTTTGTGGGCATCGCTCTCTCTCCAGATCTTTTGTGTGTGTGTGT 881
Db	879 CATTTGTGGCTGGCTCTTTCATCGGCATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 938





Qy	270	ACATCTCTGCGCTGACCGATCTGGGAGGCGCTCGACCCCTGTGTGGCTGTTTGTAGTGCTTG	322
Db	355	ACCTCACAAAGTGACCCGGATGCGATGGAATCGACCCCTGGTGTGCTCTGATGCTGG	411
Qy	330	GAGGCGTCATGTCTGCTGGCTTTGCGGCTGTGCATCGGGGCTCTCCGGAGAACTT	381
Db	415	GCGTGGTATGATTTCACTCCCTGGGTTTCGCGGCTGCGTGGGGCTCTCGCGGAGAAATATCT	471
Qy	390	TCCTGCTCAAGTTTTTCTCAGTGTTCTTGGGCTCATCTTCTTCCTGGAGCTCGCAACAG	441
Db	475	GCTTGCTCAACTTTTCTGTGGCAACATCGTGTCTATCTTCTTCTGGAGCTGGCTGTGG	531
Qy	450	GGATCTTGCGCCTTCGTATTCAAGGACTGGATTTCGAGACCAAGCTCAATTTCTTCATTAAACA	501
Db	535	CCGTGCTGCGCTTCTCTGTTCCAGGACTGGTGAGGGACCGGTTCCGGGAGTTCTTCGAGA	591
Qy	510	ACAAGCTCAAGCCCTATCGGGATGACATGTGACTCCAGAACTCATGTACTTTGCTCAGG	561
Db	595	GCAACATCAAGTCTCTACCGGGACGATATCGATCTGCAAACTCATCGACTCCCTTCAGA	651
Qy	570	AATATTGTCTTGTCTGCGGACCCGAGGCCCTAATGACTGGAACCTCAATATCTATTCA	621
Db	655	AAGCTAACAGTGTGTGGCGCATATGGCCCTGGAAGACTGGGACCTCAACGCTACTATTCA	711
Qy	630	ACTGCATCTGACTTGAAACCGGACCGAGAGCGCTGCGGGGTGCCCTTCTCTGCTGTGTCA	681
Db	715	ATTGCAGCGTGCCAGCTACAGCCGAGAGAGTGCGGGGTCCCTTCTCTGCTGTGGTGC	771
Qy	690	GGGACCTTGG- - -ATGTCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGG	741
Db	775	CAGATCTCTGCGCAAAAGTTGTGAAACACACAGTGTGGATATGATGTCAAGATTTCAGCTGA	831
Qy	746	AGCTGGAGCAGCGGCTCCATACACACCAAGGCTGTGTGGGCCAGTTTGAGAAGTGGC	801
Db	835	AGAGCAAGTGGGATGAGTCCATCTTTCAGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGC	891
Qy	806	TGCAGGACAACCTGATCGTGTGGCTGGGGTCTTTGTGGGCATCGCTCTCTCCACGATCT	861
Db	895	TCCCGCGGAACATTACATTTGCTGGCTGGCGTCTTCATCGGCATCTCGCTGTTCAGATAT	951
Qy	866	TTGTATCTGCTGCGCCAGAACCTTTGTGAGTGAATCAAGCGAGTGAAGGCC	918
Db	955	TTGCGATCTTCTTCGCAAGGACGCTGATCTCAGACATCGAGCAGTGAAGGCC	1007

Search completed: November 21, 2003, 21:46:20  
Job time : 774 secs

```

; FILE REFERENCE: 7053-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 125
; LENGTH: 2715
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-753-125

Query Match      13.9%; Score 353.4; DB 11; Length 2715;
Best Local Similarity 67.0%; Pred. No. 3.7e-92;
Matches 518; Conservative 0; Mismatches 251; Indels 4; Gaps 1

QY      150  CCGAGGTCGGCTGCTGCGGGAAATACCTTCCTGTTTGGCTTCAACATGTTTCTCGGTC 209
Db      235  CCAAGGTCAGCTGCTGGGTACAGTACCTCTTTTCAGCTCAACATCATCTTCTGTTGG 294
QY      210  TGGGAGCCCTGTTTCCTGCGCCATCGSCCTCTCGGCTGGGGTGGAGAGGGTGTCTCTCA 269
Db      295  CTGAGATTGTCTTCTTCTTGAGTCGGGCTGTGGGATGAGCGAAAGGTTGTGCTGCCG 354

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:42:48 ; Search time 21 Seconds  
(without alignment)  
469.449 Million cell updates/sec

Title: US-09-972-970-4

Perfect score: 1283

Sequence: 1 MPKXHQHFQPEVGGCGRYF.....RAPYTPKAVASLRSGCRTT 233

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	412.5	32.2	245	4	US-09-482-273-133
2	265	20.7	60	3	US-09-188-930-181
3	265	20.7	60	3	US-09-188-930-320
4	265	20.7	60	4	US-09-312-283C-181
5	265	20.7	60	4	US-09-312-283C-320
6	240	18.7	253	3	US-09-333-599-4
7	240	18.7	253	4	US-09-499-781-4
8	239	18.6	219	2	US-08-855-140-4
9	236.5	18.4	236	3	US-08-705-771-19
10	235.5	18.4	219	2	US-08-855-140-3
11	235.5	18.4	219	2	US-08-807-044-3
12	235.5	18.4	219	5	PCT-US91-04986-2
13	234	18.2	253	3	US-09-333-599-2
14	234	18.2	253	4	US-09-499-781-2
15	232	18.1	280	2	US-08-855-140-1
16	231	18.0	237	3	US-08-808-148-3
17	228	17.8	267	3	US-08-430-225A-20
18	219.5	17.1	265	2	US-08-807-044-1
19	201	15.7	227	1	US-08-254-493-1
20	201	15.7	227	2	US-08-253-751-6
21	201	15.7	227	2	US-08-453-925-6
22	201	15.7	227	4	US-08-403-253A-6
23	201	15.7	227	4	US-08-435-816A-6
24	201	15.7	228	1	US-08-408-222B-1
25	200.5	15.6	281	3	US-08-808-148-4
26	190	14.8	252	3	US-08-705-771-17
27	163	12.7	241	3	US-08-808-148-1

28	163	12.7	241	3	US-09-020-956-114	Sequence 114, App
29	163	12.7	241	3	US-09-030-607-114	Sequence 114, App
30	163	12.7	241	4	US-09-439-313-114	Sequence 114, App
31	163	12.7	241	4	US-09-352-616A-114	Sequence 114, App
32	163	12.7	241	4	US-09-232-149A-114	Sequence 114, App
33	134.5	10.5	204	4	US-09-149-476-429	Sequence 429, App
34	118.5	9.2	260	3	US-08-957-130-15	Sequence 15, Appl
35	108	8.4	258	3	US-08-957-130-13	Sequence 13, Appl
36	96.5	7.5	198	4	US-09-328-352-8198	Sequence 8198, Ap
37	94.5	7.4	101	3	US-08-905-223-443	Sequence 443, App
38	87	6.8	384	3	US-08-852-824-4	Sequence 4, Appl
39	87	6.8	384	4	US-09-731-030A-11	Sequence 11, Appl
40	87	6.8	384	4	US-09-731-030A-13	Sequence 13, Appl
41	82.5	6.4	361	4	US-09-328-352-5459	Sequence 5459, Ap
42	82.5	6.4	397	4	US-09-328-352-7457	Sequence 7457, Ap
43	81	6.3	123	4	US-09-149-476-693	Sequence 693, App
44	80.5	6.3	140	1	US-08-133-347-5	Sequence 5, Appl
45	80.5	6.3	488	4	US-09-328-352-8050	Sequence 8050, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-482-273-133  
; Sequence 133, Application US/09482273  
; Patent No. 6534631  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 71 Human Secreted Proteins  
; FILE REFERENCE: P2030P1  
; CURRENT APPLICATION NUMBER: US/09/482,273  
; EARLIER FILING DATE: 2000-01-13  
; EARLIER APPLICATION NUMBER: PCT/US99/15849  
; EARLIER FILING DATE: 1999-07-14  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/092,922  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/092,956  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 267  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 133  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (245)  
; OTHER INFORMATION: Xaa equals stop translation  
US-09-482-273-133

Query Match	32.2%	Score 412.5;	DB 4;	Length 245;
Best Local Similarity	44.2%	Pred. No. 8.7e-38;		
Matches	87;	Conservative	36;	Mismatches 59;
				Indels 15; Gaps 3;
Qy	36	LAIGLWANGKGVLSNISALTDLGGDPVWLFVVVGWVMSVLGPAGCIGALRENTFLIKF	95	
Db	2	VAVGVVRLMKHAEALACL----	AVDPAILLVGVLMFLTFCCGIGSLRENICLLQT	57
Qy	96	FSVFLGLIFLELATGILAFVFKDWIRDQNLFPINNKKAYRRDDIDQLNLIIDFAQYWSC	155	
Db	58	FSICLTAVFLQLAAGILGFVFSDKARGKVSSEINNAIVHYRDDLDLQNLIDFGQKFKSC	117	
Qy	156	CGARGPNDNLNIFYNCTDLNPSRRCGVFPSCVRDP--AMSTPSPVAMSGSNWSS	213	
Db	118	CGGISYKDSQNYFYNCSEDPNPSRCSVPSCCLPTPDQAVINTCGCGMQAFDYLEAS	177	
Qy	214	RAPYTPKAVASLRSGC	230	
Db	178	KVIYT-----NGC	185	

```
RESULT 2
US-09-188-930-181
; Sequence 181, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 181
; LENGTH: 60
; TYPE: PRT
; ORGANISM: mouse
US-09-188-930-181

Query Match      20.7%; Score 265; DB 3; Length 60;
Best Local Similarity 75.9%; Pred. No. 3.1e-22;
Matches 44; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 11 PEVCCGKYFLFGNIVFWLGAFLAIGLWAWKEGVLSNISALTDLGGDPPWLVFV 68
Db 3 PEVSCCIKYFIFGNVIFWFLGITFLGIGLWAWNEKGVLSNISITDLGGFDPWVWLF 60

RESULT 3
US-09-188-930-320
; Sequence 320, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Mouse
US-09-188-930-320

Query Match      20.7%; Score 265; DB 3; Length 60;
Best Local Similarity 75.9%; Pred. No. 3.1e-22;
Matches 44; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 11 PEVCCGKYFLFGNIVFWLGAFLAIGLWAWKEGVLSNISALTDLGGDPPWLVFV 68
Db 3 PEVSCCIKYFIFGNVIFWFLGITFLGIGLWAWNEKGVLSNISITDLGGFDPWVWLF 60

RESULT 4
US-09-312-283C-181
; Sequence 181, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orrust, Rene
; APPLICANT: Murison, James D.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/333,599
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
US-09-312-283C-181

Query Match      20.7%; Score 265; DB 4; Length 60;
Best Local Similarity 75.9%; Pred. No. 3.1e-22;
Matches 44; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 11 PEVCCGKYFLFGNIVFWLGAFLAIGLWAWKEGVLSNISALTDLGGDPPWLVFV 68
Db 3 PEVSCCIKYFIFGNVIFWFLGITFLGIGLWAWNEKGVLSNISITDLGGFDPWVWLF 60

RESULT 5
US-09-312-283C-320
; Sequence 320, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 320
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-320

Query Match      20.7%; Score 265; DB 4; Length 60;
Best Local Similarity 75.9%; Pred. No. 3.1e-22;
Matches 44; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 11 PEVCCGKYFLFGNIVFWLGAFLAIGLWAWKEGVLSNISALTDLGGDPPWLVFV 68
Db 3 PEVSCCIKYFIFGNVIFWFLGITFLGIGLWAWNEKGVLSNISITDLGGFDPWVWLF 60

RESULT 6
US-09-333-599-4
; Sequence 4, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/333,599
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
US-09-333-599-4

Query Match      20.7%; Score 265; DB 4; Length 60;
Best Local Similarity 75.9%; Pred. No. 3.1e-22;
Matches 44; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 11 PEVCCGKYFLFGNIVFWLGAFLAIGLWAWKEGVLSNISALTDLGGDPPWLVFV 68
Db 3 PEVSCCIKYFIFGNVIFWFLGITFLGIGLWAWNEKGVLSNISITDLGGFDPWVWLF 60
```

```
; SEQ ID NO 4
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-333-599-4

Query Match
Best Local Similarity 31.4%; Pred. No. 1.4e-18; Length 253;
Matches 59; Conservative 32; Mismatches 85; Indels 12; Gaps 6;

QY 8 FQEPVGC---CGKYFLGFNIVFWLGAFLAIGLWAGKGVLSNISALTDLGLDPV 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 FNEKATCGTVCLKYLFTYNCFFWLAGLAVMAGIWTALKS--DYISLLASSTYLATA 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 65 WLFVVVGWMSVLGAGCIGALRENTFLKPFVFLGLIFLELATGILAFVKDWIRDO 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 YILVAGVWMTGVLGCATFKERNLLRYFILLIIFLEIITAGILAVVYQQLNTE 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 LNFFINN--VKAYRDD--IDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSR 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 LKENLKDWTWKRYHQSGHEGVSSAVDKLQEFHC CGSNNSQDQDSEWIRSGEAD-SRV- 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 CGVPFSCC 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 --VPDSCC 185

RESULT 7
US-09-499-781-4
; Sequence 4, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/499,781
; CURRENT FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 09/333,599
; PRIOR FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-499-781-4

Query Match
Best Local Similarity 31.4%; Pred. No. 1.4e-18; Length 253;
Matches 59; Conservative 32; Mismatches 85; Indels 12; Gaps 6;

QY 8 FQEPVGC---CGKYFLGFNIVFWLGAFLAIGLWAGKGVLSNISALTDLGLDPV 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 FNEKATCGTVCLKYLFTYNCFFWLAGLAVMAGIWTALKS--DYISLLASSTYLATA 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 65 WLFVVVGWMSVLGAGCIGALRENTFLKPFVFLGLIFLELATGILAFVKDWIRDO 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 YILVAGVWMTGVLGCATFKERNLLRYFILLIIFLEIITAGILAVVYQQLNTE 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 LNFFINN--VKAYRDD--IDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSR 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 LKENLKDWTWKRYHQSGHEGVSSAVDKLQEFHC CGSNNSQDQDSEWIRSGEAD-SRV- 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 CGVPFSCC 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 --VPDSCC 185
```

RESULT 8  
US-08-855-140-4

```
; Sequence 4, Application US/08855140
; Patent No. 5854022
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,140
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0296 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1279546
; US-08-855-140-4

Query Match
Best Local Similarity 30.9%; Pred. No. 1.4e-18; Length 219;
Matches 63; Conservative 29; Mismatches 72; Indels 40; Gaps 7;

QY 18 KYFLGFNIVFWLGAFLAIGLW--AWGKGVLSNISALTDLGLDPVWLVVVVGV 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 KYVLPFNLLFWVCGCCILGFIYLVQNTVGVLFRLNLPFLT-LGN-----ILVIGSII 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 75 SVLGFACIGALRENTFLKPFVFLGLIFLELATGILAFVKDWIRDOINFFINN 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 MWAFGLCMGSIKENKCLMSFFVLLLIILLAEVTIALLFVYBOKLNTLVAEGLN 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 135 AYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRRCGVFPSCC 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 HYISDNTMKAWDFIQQLQCCGVNGSSDMT-----SGPPSSC----- 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 MSSTPSVAMMSG-----SNMWS 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 ----PSGADVQGCYNKAKSWFHSN 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-08-705-771-19
; Sequence 19, Application US/08705771
; Patent No. 6054289
; GENERAL INFORMATION:
```

```

; APPLICANT: Paul Moore, Reiner Gentz, Hongjin Ji,
; APPLICANT: Jian Ni and Jing-Shan Hu
; TITLE OF INVENTION: Human Genes, Sequences and
; TITLE OF INVENTION: Expression Products
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/705,771
; FILING DATE: August 30, 1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-346 (PF196)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-705-771-19

Query Match 18.4%; Score 236.5; DB 3; Length 236;
Best Local Similarity 28.9%; Pred. No. 3e-18;
Matches 56; Conservative 34; Mismatches 83; Indels 21; Gaps 6;

Qy 18 KYFLFGNIVFWVLGALFLATGLWAWGKGVLSNISALTDLGGLDPVW---LFVVVGVM 74
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
10 KYLFAFNLLFWLGGCVGLGVGIMLAQAQGSFATLSS-----SFFSLWAANLLIITGAFV 64
Qy 75 SVLGFAGCIGALRENTLLKFFSVFLGLIFLELATGILAFVFXDWIRDLQNFNNVK 134
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
65 MAIGFVGCIGAIKENKCLLTFLLLLLVFLLEGTIALFFAYTDKIDRYAQDLKXGLH 124
Qy 135 AY--RDDIDLQNLIDFAQYWSCCGARGPNDNLNIYFNCTDLNPSRRCGVFPSCVVD 192
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
125 LYGTQGNVGLTNWSIIQTDFRCGVSNYTDM-FEVY-----NATR-----VPDSCCLEF 173
Qy 193 PAMSTPVSVMMSG 206
Db :|::|
174 SESCGLHAPALVEG 187

RESULT 10
US-08-855-140-3
; Sequence 3, Application US/08855140
; Patent No. 5854022
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto

```

```

; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,140
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0296 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 180141
; US-08-855-140-3

Query Match 18.4%; Score 235.5; DB 2; Length 219;
Best Local Similarity 31.0%; Pred. No. 3.5e-18;
Matches 54; Conservative 31; Mismatches 64; Indels 25; Gaps 5;

Qy 18 KYFLFGNIVFWVLGALFLAIGLM--AWGKGVLSNISALTDLGGLDPVWLVVVVGVM 74
Lj |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
10 KYVLFFENLLFWIGCCILGFIYLLIHNPNGLVFLHNLPSLT-LGN-----VFVIVGSI 63
Qy 75 SVLGFAGCIGALRENTLLKFFSVFLGLIFLELATGILAFVFXDWIRDLQNFNNVK 134
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
64 MVAFGLGCMGSIKENKCLLMSFFILLIILAEVTLAILLFEVYBOKLNEYVAKGLTDSIH 123
Qy 135 AYRDDIDLQNLIDFAQYWSCCGARGPNDNLNIYFNCTDLNPSRRCGVFPSC 188
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
124 RYHSDNSTKAWDISIQSFLQCCGNGTSDMT-----SGPPASC 161

RESULT 11
US-08-807-044-3
; Sequence 3, Application US/08807044
; Patent No. 5863735
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:

```

```

? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/07/553,759
? FILING DATE: 13-JUL-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/498,809
? FILING DATE: 23-MAR-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/379,076
? FILING DATE: 13-JUL-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/160,416
? FILING DATE: 25-FEB-1988
? ATTORNEY/AGENT INFORMATION:
? NAME: Wall, Margaret M.
? REGISTRATION NUMBER: 33,462
? REFERENCE/DOCKET NUMBER: 11-88C
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 303-499-8080
? TELEFAX: 303-499-8089
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 219 amino acids
? TYPE: AMINO ACID
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? PCT-US91-04986-2

Query Match 18.4%; Score 235.5; DB 5; Length 219;
Best Local Similarity 31.0%; Pred. No. 3.5e-18;
Matches 54; Conservative 31; Mismatches 64; Indels 25; Gaps 5;

Qy 18 KYFLFGFNIVFWVLGALFLAIGLW--AWGEKGVL-SNISALTDLGGLDPWVLFWVVGVM 74
Db 10 KYVLFNLLFWICGCIIGFGYLIHNHFGVLFHNLPSLT-LGN-----VPVIVGSII 63

Qy 75 SVLGPAAGCIGALRENTFLKFPFVFLGLIPFLELATGILAFVKDWIRDLQINFFINNVK 134
Db 64 MVVAFLGCMGSIKENKCLLMSPFILLIILLAEVTLAILLFVVEQKLVYVAKGLTDSIH 123

Qy 135 AYRDDIDLQNLIDFAQYHSCCGARGPDWNLNIYFNCTDLNFSRRCGVFFSC 188
Db 124 RYHSDNSTKAWDSDIOSFLQCCGINTGSDWT-----SGPPASC 161

```

```

Db      124 RYHSDNSTKAAWDSIQSFLOCCGNGTSDWT-----SGFPASC 161

RESULT 13
US-09-333-599-2
; Sequence 2, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/333,599
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-333-599-2

Query Match      18.2%; Score 234; DB 3; Length 253;
Best Local Similarity 31.0%; Pred. No. 6,3e-18;
Matches 61; Conservative 29; Mismatches 77; Indels 30; Gaps 7;

Qy      8 FQEPVEGC---CGKTYFLPGFNIVFWVLGALFLAIGLWANGCKGVLSNISALTDLGGDPV 64
Db      4 FNEKKTTCGTGCLKYLFTYNCFCFVLGAGLVAVMGVITTLALKS--DYISLLASGTYLATA 61

```

Qy 65 WLFVVVGVMVSLGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDWIRQ 124  
Db 62 YILVAVGTVMVTVGLGCGCATFKERNLLRLYFILLIIFLEITAGILAYAY----YQQ 117  
Qy 125 LNFFINNKKAVRDDID-----LQNLIDFAQYVWSCGARGPNDNLIYFNCTD 174  
Db 118 LNTELKENLK---DTMTKRYHQPGEAVTSVDQLQEFHCCGNSQDWRDSEWIR--- 171  
Qy 175 LNPSRRCG--VPFSCC 189  
Db 172 ---SQEAGRVVPDSCC 185

## RESULT 14

US-09-499-781-2  
; Sequence 2, Application US/09499781  
; Patent No. 6498014  
; GENERAL INFORMATION:  
; APPLICANT: Testa, Jacqueline E.  
; APPLICANT: Quigley, James P.  
; APPLICANT: Seandel, Marco  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS  
; TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS  
; FILE REFERENCE: SUNY  
; CURRENT APPLICATION NUMBER: US/09/499,781  
; CURRENT FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 09/333,599  
; PRIOR FILING DATE: 1999-06-15  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 253  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-499-781-2

Query Match 18.2%; Score 234; DB 4; Length 253;  
Best Local Similarity 31.0%; Pred. No. 6.3e-18;  
Matches 61; Conservative 29; Mismatches 77; Indels 30; Gaps 7;  
Qy 8 FOEPEVGC---CGKYFLFGFNIVFVLGALFLAIGLWAWGKGVLSNISALTDLGLDPV 64  
Db 4 FNEKTKCTGTVCLKYLLFTYNCCEFWLAGLVAVGIWTLAKS--DYISLASGYLATA 61  
Qy 65 WLFVVVGVMVSLGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDWIRQ 124  
Db 62 YILVAVGTVMVTVGLGCGCATFKERNLLRLYFILLIIFLEITAGILAYAY----YQQ 117  
Qy 125 LNFFINNKKAVRDDID-----LQNLIDFAQYVWSCGARGPNDNLIYFNCTD 174  
Db 118 LNTELKENLK---DTMTKRYHQPGEAVTSVDQLQEFHCCGNSQDWRDSEWIR--- 171  
Qy 175 LNPSRRCG--VPFSCC 189  
Db 172 ---SQEAGRVVPDSCC 185

## RESULT 15

US-08-855-140-1  
; Sequence 1, Application US/08855140  
; Patent No. 5854022  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Guegler, Karl J.  
; TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto

; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/855,140  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0296 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 280 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: MYOMN001  
; CLONE: 779308  
; US-08-855-140-1

Query Match 18.1%; Score 232; DB 2; Length 280;  
Best Local Similarity 27.5%; Pred. No. 1.2e-17;  
Matches 60; Conservative 30; Mismatches 92; Indels 36; Gaps 6;  
Qy 15 CCGKYFLFGFNIVFVLGALFLAIGLWAWGKGVLSNISALTDLGLDPVWLFVVVGVM 74  
Db 7 CCLKYMFLFNLIFFWLCGCGLLGVWLSVQGNFATFS--PSFPSSLAAANLVIAIGTIV 64  
Qy 75 SVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDWIRQDLNFFINNKK 134  
Db 65 MVTGFLGCLGAIKENKCLLSFFIVLLVTLAEILILFILFFVYMDKYNENAKDKLEGL 124  
Qy 135 AY--RDDIDLQNLIDFAQYVWSCGARGPNDNLIYFNCTDLNPSRRCGVPFSCCVRD 192  
Db 125 LYHTENNVLKNAWNIQAEMRCGVTDYTDW-----YFVLGENTVPDRCCM-- 171  
Qy 193 PAMSTPSVAMSGSNWSWSRAPYTPKAVWASLRSGC 230  
Db 172 -----ENSGGGRNGTTP--LW---RTGC 190

Search completed: November 21, 2003, 13:46:00  
Job time : 22 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:41:03 ; Search time 35 Seconds  
(without alignments)  
1717.893 Million cell updates/sec

Title: US-09-972-970-4  
Perfect score: 1283  
Sequence: 1 MPKGKHQFQPEVCGCKYF.....RAPVTPKAVASLRSGCRTT 233

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_todent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1066	83.1	263	4 Q96FV3	Q96fv3 homo sapien
2	1048	81.7	270	11 Q91V16	Q91v16 mus musculu
3	1048	81.7	270	11 Q9D7W4	Q9d7w4 mus musculu
4	885	69.0	266	11 Q9CYT0	Q9cyt0 mus musculu
5	646.5	50.4	270	11 Q8QZY6	Q8qzy6 mus musculu
6	638.5	49.8	270	4 Q8NG11	Q8ng11 homo sapien
7	635.5	49.5	270	4 Q9H0U1	Q9h0u1 homo sapien
8	554	43.2	253	4 Q9BU34	Q9bu34 homo sapien
9	487.5	38.0	123	4 Q9UKB9	Q9ukb9 homo sapien
10	482.5	37.6	269	5 Q9VMJ6	Q9vmj6 drosophila
11	472.5	36.8	283	11 Q8R3S2	Q8r3s2 mus musculu
12	423	33.0	188	4 Q96S98	Q96s98 homo sapien
13	421.5	32.9	291	5 Q9VG3	Q9vg3 drosophila
14	419.5	32.7	308	5 Q22495	Q22495 caenorhabdi
15	412.5	32.2	274	5 Q9NB16	Q9nb16 drosophila
16	411.5	32.1	304	5 Q9W4X6	Q9w4x6 drosophila

17	371.5	29.0	302	5 Q9U1L0	Q9u1l0 drosophila
18	336.5	26.2	206	11 Q99L35	Q99l35 mus musculu
19	308.5	24.0	355	4 Q8N548	Q8n548 homo sapien
20	308.5	24.0	355	4 Q9H1Z9	Q9h1z9 homo sapien
21	283	22.1	427	5 Q44582	Q44582 caenorhabdi
22	283	22.1	451	5 Q95X63	Q95x63 caenorhabdi
23	269	21.0	248	4 Q96S38	Q96s38 homo sapien
24	269	21.0	267	4 Q8WU1	Q8wu1 homo sapien
25	265.5	20.7	331	11 Q8VCFS	Q8vcfs mus musculu
26	259	20.2	349	4 Q9H7Q1	Q9h7q1 homo sapien
27	251	19.6	236	4 Q8NB19	Q8nb19 homo sapien
28	242	18.9	225	4 Q8N8E5	Q8n8e5 homo sapien
29	233.5	18.7	245	11 Q99L36	Q99l36 mus musculu
30	239	18.6	253	11 Q92L37	Q92l37 mus musculu
31	235.5	18.4	253	11 Q9D1D1	Q9d1d1 mus musculu
32	235	18.3	239	11 Q8BJU2	Q8bjuz2 mus musculu
33	228.5	17.8	235	11 Q8R3G9	Q8r3g9 mus musculu
34	228.5	17.8	239	5 Q9NB10	Q9nb10 manduca sex
35	228.5	17.8	247	13 Q9PT80	Q9pte0 gallus gall
36	222.5	17.3	291	5 Q9NB09	Q9nb09 manduca sex
37	222	17.3	224	13 Q9IBC9	Q9ibc9 gallus gall
38	222	17.3	248	5 Q9VLH1	Q9v1h1 drosophila
39	220	17.1	249	11 Q8BT06	Q8bt06 mus musculu
40	219.5	17.1	235	11 Q55158	Q55158 rattus norv
41	218.5	17.0	267	5 Q9V3E5	Q9v3e5 drosophila
42	214	16.7	227	5 Q8T9S4	Q8t9s4 dermacentor
43	214	16.7	236	5 Q9NB08	Q9nb08 manduca sex
44	210	16.4	236	11 Q91V78	Q91v78 mus musculu
45	209	16.3	245	5 Q9VID1	Q9vid1 drosophila

## ALIGNMENTS

### RESULT 1

Q96FV3 PRELIMINARY; PRT; 263 AA.  
 ID Q96FV3;  
 AC Q96FV3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Similar to RIKEN CDNA 2210021G21 gene.  
 OS Homo sapiens (Human).  
 UC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_taxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC010405; AAH10405.1; -.  
 DR InterPro; IPR00301; Transmem 4.  
 DR Pfam; PF00335; transmembrane4; 1.  
 DR PRINTS; PR00259; TWFOUR.  
 SQ SEQUENCE 263 AA; 29569 MW; 820189FB903DDB14 CRC64;

Query Match	83.1%	Score 1066;	DB 4;	Length 263;
Best Local Similarity	99.5%	Pred. No. 5.4e-83;		
Matches 193;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MPKGKHQFQPEVCGCKYFLFGFNIVFWLGLFLAIGLWAWGKGVLSNISALTDLGG	60	
Db	1	MPKGKHQFQPEVCGCKYFLFGFNIVFWLGLFLAIGLWAWGKGVLSNISALTDLGG	60	
Qy	61	LDPVWLVVVGWVMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW	120	
Db	61	LDPVWLVVVGWVMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVFKDW	120	
Qy	121	IRQLNFFINNKKAYRDDIDLQNLIDFAQEYNSCCGARGNDWNLNIYFNCTDLNFSRE	180	
Db	121	IRQLNFFINNKKAYRDDIDLQNLIDFAQEYNSCCGARGNDWNLNIYFNCTDLNFSRE	180	

	Qy	181	RCGVPFSCCVDPDA	194       
	Dd	181	RCGVPFSCCVDRPA	194       
RESULT 2				
ID	Q91VI6	PRELIMINARY;		PRT;    270 AA.
AC	Q91VI6;			
DT	01-DEC-2001	(TrEMBLrel. 19,	Created)	
DD	01-DEC-2001	(TrEMBLrel. 19,	Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23,	Last annotation update)	
DE	RIKEN CDNA 2210021G21 gene.			
GN	2210021G2IRIK.			
GZ	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Breast tumor;			
RA	Straussberg R.;			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; BC010346; AALH0346.1; -			
MGD	NCBI_1921507; 2210021G2IRik.			
DR	InterPro; IPR000301; Transmem 4.			
DR	pfam; PF00335; transmembrane4; 1.			
DR	PRINTS; PR00259; TMFOUR.			
SD	SEQUENCE      270 AA;     30125 MW;    660FE581CIEA42DA CRC64;			
Query Match                  81.7%; Score 1048; DB 11; Length 270;				
Best Local Similarity        97.4%; Pred. No. 1.9e-81;				
Matches 189; Conservative    2; Mismatches    3; Indels    0; Gaps    0;				
QY	1	MPGHKHOFORPEVCGCKGYFLFGFNIVFVLWGALFLAIGLWAWGEKGVLSNISALTDLGG	60	
DB	1	MPGHKHOFORPEVCGCKGYFLFGFNIVFVLWGALFLAIGLWAWGEKGVLSNISALTDLGG	60	
QY	61	LDPWLFWVGVMNVLGFGAGICGALRENTLLKFPSVFGLGIFFLELATGILAFVKDW	120	
DB	61	LDPWLFWVGVMNVLGFGAGICGALRENTLLKFPSVFGLGIFFLELATGILAFVKDW	120	
QY	121	IROQLNFIINNKKAYRDDLDLOLPDAQBWSGCCARGNDNLNIYNCTDLNPFSR	180	
DB	121	IROQLNFIINNKKAYRDDLDLOLPDAQBWSGCCARGNDNLNIYNCTDLNPFSR	180	
QY	181	RCGVPFSCCVDRPA	194	
DB	181	RCGVPFSCCVDRPA	194	
RESULT 3				
ID	Q9D7M4	PRELIMINARY;		PRT;    270 AA.
AC	Q9D7M4;			
DT	01-JUN-2001	(TrEMBLrel. 17,	Created)	
DD	01-JUN-2001	(TrEMBLrel. 17,	Last sequence update)	
DT	01-JUN-2001	(TrEMBLrel. 17,	Last annotation update)	
DE	2210021G2IRik protein.			
GN	2210021G2IRIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Stomach;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino H., Adachi J., Fukuda S.,			
RA	Arakawa T., Haga A., Fukuishi Y., Konno H., Adachi J., Yamanaaka I.,			
RA	Aizawa K., Izawa M., Niehi K., Kiyoosawa H., Kondo S., Yamanaaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasakura T., Saito R.,			
RA	Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			

```

RESULT 5
QBQZY6
ID QBQZY6 PRELIMINARY; PRT; 270 AA.
AC QBQZY6;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Similar to region containing hypothetical protein MGCI1352, slingshot
DE (Hypothetical protein) (Similar to transmembrane 4 superfamily member
DE 9 homolog).
DE D14ERTD226E OR AA958793.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP RP
RC TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP RP
RC Strausberg R.;
RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP RP
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RT Nature 420:563-573 (2002).
DR EMBL; BC024611; AAH24611.1; -
DR EMBL; BC025568; AAH25568.1; -
DR EMBL; BC026574; AAH26574.1; -
DR EMBL; AK030593; BAC27035.1; -
DR MGI; 1196325; D14ERTD226E.

```

Query Match	49.8%;	Score 638.5;	DB 4;	Length 270;
Best Local Similarity	49.4%;	Pred. No. 1.4e-46;		
Matches 119; Conservative 37; Mismatches 66; Indels 19; Gaps 2;				
QY	5	HQHFOPEYGGCKGYFLGFINVFVLGLAFLAIGLWAMEKGVLISNLSALTDLGLDIPV	64	
Db	:	: : :	:	:
QY	3	YYRSNAKVCWKYLLFSYNIFLWAGVVFGVLGWAMSEKGVSDLTKVTMRMHIDIPV	62	
Db	:	: : :	:	:
QY	65	WLUFVVVGVMVSLGFAGCTGALRENTFLFKPFSVFLGLIFFLELATGIILAFYFMDIRIQ	124	
Db	:	: : :	:	:
QY	63	VLVLMVGVMVFELGFAGCVGALRENICLINFPFCGTIVLIFFLELAVALFAFLFQDMVRDR	122	
Db	:	: : :	:	:
QY	125	LNFFINNKKAYRDDIDLQNLDIFAQYMWSCCGARGPNDDNNINVENCTDLNPSRCCGV	184	
Db	:	: : :	:	:
QY	123	PREPFESNIKSRYDDIDLQNLDISLOKANQCAGYPEDWDUNVFNCSGASYSREKCGV	182	
Db	:	: : :	:	:
QY	195	PFSCCVRDPMSSTPS-----VMMMSGNW-----SWSSRAPYTPKAVMAS	225	
Db	:	: : :	:	:
QY	193	PFSCCVPDPQAQRVNVTQCYDVRIQLKSKWDESIPTKGCIQIALESWLPENIYIVAGVFIA	242	
Db	:	: : :	:	:
QY	226	L 226		

```
DB 243 I 243

RESULT 7
Q9H0U1 PRELIMINARY; PRT; 270 AA.
AC Q9H0U1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
GN DKFZP564B1037.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecker M., Blocker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Pousetka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435 (2001).
DR EMBL; AL136638; CAB66573.1; -.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 30696 MW; 45589912DC5381108 CRC64;

Query Match 49.5%; Score 635.5; DB 4; Length 270;
Best Local Similarity 49.0%; Pred. No. 2.5e-46;
Matches 118; Conservative 38; Mismatches 66; Indels 19; Gaps 2;

QY 5 HQHFOEPEVGGCGKYFLFGFNIVFVLGALFLAIGLWAMGKGVLSNISALTDLGLDPV 64
DB 3 YVYSNAKVCWYKYLFSYNIIFW-----GVLSDLTKVTRMHGIDPV 62
QY 65 WLFFVVGVMVTLGPGAGCGALRENTFLKFFSVFLGIFFLELATGILAFVKDWIRDO 124
DB 63 VLVLMGVVMPTLGPAGCGVGLRENICLNFPCGTVILIFFLELAVAVLFLQDWVRDR 122
QY 125 LNFFINNKKAYRDDIDLQNLIDPAQYWSCCGARGPNDNLIYFNCTDLNPSRRCGV 184
DB 123 PREFPESNIKSYRDDIDLQNLIDSLQKANCQCGAYGPDWDNLVYFNCSGASYREKCGV 182
QY 185 PFSCCVRDPAMSSSTPS-----VAMMSGNW-----SWSSRAPYTPKAVNAS 225
DB 183 PFSCCVPDPAQKVNTQCGYDVRITQKSKWDESIPTKCIQALSWLPNIIYVAGVFIA 242
QY 226 L 226
DB 243 I 243

RESULT 8
Q9BU34 PRELIMINARY; PRT; 253 AA.
AC Q9BU34;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Similar to transmembrane 4 superfamily member 9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

Query Match 38.0%; Score 487.5; DB 4; Length 123;
Best Local Similarity 85.0%; Pred. No. 4.3e-34;
Matches 96; Conservative 1; Mismatches 11; Indels 5; Gaps 2;

QY 1 MPGKHQHFQEPVEGCCGKYFLFGFNIVFVLGALFLAIGLWAMGKGVLSNISALTDLGG 60
DB 13 MPGKHQHFQEPVEGCCGKYFLFGFNIVFVLGALFLAIGLWAMGKGVLSNISALTDLGG 72
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002920; AA02920.1; -.
DR InterPro; IPR00301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
KW Transmembrane.
SQ SEQUENCE 253 AA; 28876 MW; F83140442CFCAC16 CRC64;

Query Match 43.2%; Score 554; DB 4; Length 253;
Best Local Similarity 45.2%; Pred. No. 2e-39;
Matches 109; Conservative 34; Mismatches 62; Indels 36; Gaps 3;

QY 5 HQHFOEPEVGGCGKYFLFGFNIVFVLGALFLAIGLWAMGKGVLSNISALTDLGLDPV 64
DB 3 YVYSNAKVCWYKYLFSYNIIFW-----GVLSDLTKVTRMHGIDPV 45
QY 65 WLFFVVGVMVTLGPGAGCGALRENTFLKFFSVFLGIFFLELATGILAFVKDWIRDO 124
DB 46 VLVLMGVVMPTLGPAGCGVGLRENICLNFPCGTVILIFFLELAVAVLFLQDWVRDR 105
QY 125 LNFFINNKKAYRDDIDLQNLIDPAQYWSCCGARGPNDNLIYFNCTDLNPSRRCGV 184
DB 106 PREFPESNIKSYRDDIDLQNLIDSLQKANCQCGAYGPDWDNLVYFNCSGASYREKCGV 165
QY 185 PFSCCVRDPAMSSSTPS-----VAMMSGNW-----SWSSRAPYTPKAVNAS 225
DB 166 PFSCCVPDPAQKVNTQCGYDVRITQKSKWDESIPTKCIQALSWLPNIIYVAGVFIA 225
QY 226 L 226
DB 226 I 226

RESULT 9
Q9UKB9 PRELIMINARY; PRT; 123 AA.
AC Q9UKB9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE F-box protein FBX23 (Fragment).
GN FBX23.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2003060; PubMed=10531035;
RA Cenciarelli C., Chiaux D.S., Guardavaccaro D., Parks W., Vidal M.,
RA Pagano M.;
RT "Identification of a family of human F-box proteins.";
RL Curr. Biol. 9:1177-1179 (1999).
DR EMBL; AF174603; AA04524.1; -.
DR Genew; HGNC:13594; FBXO23.
DR InterPro; IPR00301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR NON_TER 1
DR NON_TER 1
FT NON_TER 123
FT NON_TER 123
SQ SEQUENCE 123 AA; 13531 MW; 0A6CDB2AF450D8EF CRC64;

Query Match 38.0%; Score 487.5; DB 4; Length 123;
Best Local Similarity 85.0%; Pred. No. 4.3e-34;
Matches 96; Conservative 1; Mismatches 11; Indels 5; Gaps 2;

QY 1 MPGKHQHFQEPVEGCCGKYFLFGFNIVFVLGALFLAIGLWAMGKGVLSNISALTDLGG 60
DB 13 MPGKHQHFQEPVEGCCGKYFLFGFNIVFVLGALFLAIGLWAMGKGVLSNISALTDLGG 72
```

```

QY 61 LDPVWLVVVGVGMSVLFAGC---IGALRENTFLKFFSVFLGLIFFLELAT 110
DB 73 LDPVWLV--VCGSWRRHVAGLWAAIGALRENTFLKFFXFXFLGLIFFLELAT 123

RESULT 10
Q9VMJ6 PRELIMINARY; PRT; 269 AA.
AC Q9VMJ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CG9093 protein.
GN TSP26A OR CG9093.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Fianknoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003612; AAF52320.1; --
DR FlyBase; FBgn0031760; Tsp26A.
DR InterPro; IPR00301; Transmem. 4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; THFOUR.
SQ SEQUENCE 269 AA; 30235 MW; A05A188B3224A165 CRC64;

Query Match 37.6%; Score 482.5; DB 5; Length 269;
Best Local Similarity 46.2%; Pred. No. 2.6e-33;
Matches 92; Conservative 35; Mismatches 65; Indels 7; Gaps 5;

QY 1 MPKGKHQHF--QEPVCGCGKYFLFGFNIVFWLGVLFALGLWAGKGVLSNLSALTDL 58

```

```

DB 1 MPAAVRKRRRTSEISCCCLKYLLFASNVILWLSALLVSVGIWAWSEKGMFRNIARLHI 60
QY 59 GGLDPVWLVVVGVGMSVLFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVK 118
DB 61 -ALDPAFVLIILGGVTFLLGFMGSGALRENTCLLGAVAFVLSVILLIAEIGCAVAFVK 119
QY 119 D--WIRDQLNFFNNNVKAYRDDIDLONLIDPAQBYW-SCCGARGPNDNLNIYFNCTDL 175
DB 120 DKGWTKDQATEGLKAFIRHYREDADQONLIDWILQCCGIDGPKOWDSNNYFNCSII 179
QY 176 N-PSRRCGVFPSCCVRDP 193
DB 180 AIGSREACGVFPSCCRRRP 198

RESULT 11
Q9R3S2 PRELIMINARY; PRT; 283 AA.
AC Q9R3S2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to transmembrane 4 superfamily member (Tetraspan NET-7).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Straubeberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024685; AAH24685.1; -.
DR InterPro; IPR000301; Transmem. 4.
KW Transmembrane.
SQ SEQUENCE 283 AA; 31550 MW; 9AF1515D5B0BF5CF CRC64;

Query Match 36.8%; Score 472.5; DB 11; Length 283;
Best Local Similarity 43.5%; Pred. No. 1.9e-32;
Matches 100; Conservative 41; Mismatches 70; Indels 19; Gaps 4;

QY 3 GKHFQHFQEPVCGCGKYFLFGFNIVFWLGVLFALGLWAGKGVLSNLSALTDLGLD 62
DB 12 GDEFSFVSPLV---KYLFFFNLFVWISVMVAVGVYARLMKJAEALACL----AVD 63
QY 63 PVWLVVVGVGMSVLFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWTR 122
DB 64 PAILLIVGVLMFLTTCGCGISLRNICLLQTSCLTIVFLQLAAGILGFVFSKAR 123
QY 123 DQLNFFNNNVKAYRDDIDLONLIDPAQBYWSCCGARGPNDNLNIYFNCTDLNPSRRC 182
DB 124 GKVSEIINNAIVHYRDDLDLQNLIDFGKKFSCCGISYRDWSQWYFNCSNEDNPSRRC 183
QY 183 GVPSCCVRDP--AMSTPVSVMGSGNWSWSSRAPYTPKAVWASLRSGC 230
DB 184 SVPYSCCLPTNQAVINTMCGQGMQALDYLEASKVIYT-----NGC 224

RESULT 12
Q96S98 PRELIMINARY; PRT; 188 AA.
AC Q96S98;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tetraspan protein SB134.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang W., Li N., Wan T., Cao X.;

```

RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of <i>Drosophila melanogaster</i> "; RL Science 287:2185-2195(2000).
DR	EMBL; AE003688; AAF54570.1; -. DR FlyBase; FBgn0037848; Tep86D.
DR	InterPro; IPR000301; Transmem 4.
DR	pfam; PF00335; transmembrane4; 1.
DR	PRINTS; PR00259; TMFOUR.
SQ	SEQUENCE 291 AA; 33213 MW; E57FC55688B65C3D CRC64;
Query Match	32.9%; Score 421.5; DB 5; Length 291;
Best Local Similarity	40.8%; Pred. No. 4.4e-28;
Matches 97; Conservative	39; Mismatches 71; Indels 31; Gaps 10
Qy	7 HFQPEVGGCGKFLFGFNIVFWVLGALFLAIGLMWAGEKGVLSNLSALTDLGG--LDP 63
Db	23 HFS--YVSSCVKTMIFLNFLFWFGLLLAIGVYAFMDK-----LMDGNGLRLDT 72
Qy	64 VW-----LFVVVGVSMLVFAGCGIGALRENTFLPKFFSVFLGIFFLELATGILAF 115
Db	73 IYDVIFNISLVMIIAIGVIVFTSVFAGCLGALRENTWLKLYSMCLLFFFILEMSLAITCF 132
Qy	116 VFQDWTRDQLNF-FINNWKAYRDDLDLQNLDPAQEYWSCGA--RGPNDWNLIYFNC 172
Db	133 VPQQYNMFLEYOFTDKIIHSYRDDSLOFNIDFAQQEFNCCGSLNAGYQDMSKNEYFNC 192
Qy	173 TDLPNSRRCGVFPFCCVRDPMASSTPVSVMGWSNNWSWRAPYTPKAVMWASRSG 230
Db	193 S--SPSVERGVPSGCCINATDISS-GLVNIMCGYGVOVRSVA--ASKRIWT---SGC 243
RESULT 14	
Q22495	PRELIMINARY; PRT; 308 AA.
ID	Q22495
AC	AQ
DT	01-NOV-1996 (TrEMBLrel_01, Created)
DT	01-NOV-1996 (TrEMBLrel_01, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel_23, Last annotation update)
DE	T14G10.6 protein.
DN	T14G10.6.
GN	Cenorhabditis elegans.
OS	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Paloderinae; Caenorhabditis.
OX	NCB1_TaxID=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Wild A.;
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RL	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99069613; PubMed=9851916;
RA	none;
RT	"Genome sequence of the nematode <i>C.elegans</i> : A platform for investigating biology."
RT	Science 282:2012-2018(1998).
DR	EMBL; Z68880; CAA93092.1; -. WormPep; T14G10.6; CE06452.
DR	InterPro; IPR000301; Transmem 4.
DR	Pfam; PF00335; transmembrane4; 1.
DR	PRINTS; PR00259; TMFOUR.
SQ	SEQUENCE 308 AA; 34658 MW; E37221DD2ADE78E CRC64;
Query Match	32.7%; Score 419.5; DB 5; Length 308;
Best Local Similarity	42.9%; Pred. No. 6.9e-28;
Matches 78; Conservative	32; Mismatches 71; Indels 1; Gaps 1;
Qy	10 EPEVGGCKGYKFLFGFNIVFWVLGALFLAIGLMWAGEKGVLSNLSALTDLGGLDPVWLVFV 69
Db	31 ESEISCCVKYSVPFNVFIPLGGLLGVAIQEKYTFVWMLSKAKLYLDPTWPLLI 90

Query Match	32.2%	Score 412.5;	DB 5;	Length 274;
Best Local Similarity	41.3%;	Pred. No. 2.4e-27;		
Matches	97;	Conservative 33;	Mismatches 70;	Indels 35; Gaps 9

  

QY	13	VGCGKGYFLFGFNIVFVWLGALFLAIGLWAMGEK-----	GVLSNISAL	TDL	58
Db	9	VSQCVKYMIFLLNFVFLFCGLLLIGIYGYAFDRKWE	DANGSVRL	ENFDVFNISL	VMIL 68
QY	59	GGLDPPWFLFVVVGGVMSVLFGACIGIGALRENTFL	LKFFSVFLG	LIFFLELATG	IAFVFK 118
Db	69	AGTD---IFLV-----	SFGCVG	ALRENTFL	LKFFSYMCLLFFLLEMAIAIVCFVCP 117
QY	119	DTIRDOL-NPFFINNKKAYRDDIDQLNLDIPAQYWS	CCGGA--	RGPNDN	NLINIVFNCTDL 175
Db	118	QYMNTLEKQFTHKIHSYRDDPDLQNFIDPAQQE	FKCCGLNS	SGYQDMSK	NEVFNCS-- 175
QY	176	NPSRECGVPFSCVDRPDMSSPTSVAMGSGNWS	WSR	APYTPK	VAWSLRSGC 230
Db	176	SPSEVKCGVPYSCINATDTSI--GLVNIMCGY	GVQV-NAP	PEATKL	TWT SG 225

Search completed: November 21, 2003, 13:44:54



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:41:43 ; Search time 20 Seconds  
(without alignments)  
1120.365 Million cell updates/sec

Title: US-09-972-970-4  
Perfect score: 1283  
Sequence: 1 MPKGHQHQPPEVCGCKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	852	66.4	264	2 A59261	tetraspan TSPAN-5
2	419.5	32.7	308	2 T24912	hypothetical prote
3	283	22.1	427	2 T32652	hypothetical prote
4	243	18.9	238	2 A59265	tetraspan TSPAN-4
5	240.5	18.7	245	2 A59258	tetraspan TSPAN-6
6	240	18.7	244	1 I39368	T-cell acute lymph
7	239.5	18.7	245	2 A59260	tetraspan TSPAN-6
8	235.5	18.4	219	1 A37243	hemopoietic cell s
9	233.5	18.2	219	1 A39574	leukocyte antigen
10	231	18.0	237	1 A36056	tumor-associated a
11	228	17.8	267	1 A46493	metastasis suppres
12	220	17.1	238	1 S43511	CD63/ME491 antigen
13	217	16.9	266	2 I49561	C33/R2/IA4 - mouse
14	214	16.7	238	1 I38016	melanoma-associate
15	214	16.7	238	1 JC2397	CD63 antigen - rab
16	211	16.4	238	1 A46508	cell surface prote
17	210	16.4	236	1 A46472	antigen - mouse
18	208	16.2	226	2 I49589	cell surface prote
19	207	16.1	236	1 A35649	CD9 antigen - rat
20	206	16.1	236	1 S39262	hypothetical prote
21	205	16.1	282	2 T21696	CD9 antigen - gre
22	202	15.7	228	1 A42829	CD9 antigen [valid
23	201	15.7	228	1 A40402	cell surface glyco
24	200.5	15.6	281	1 B47629	23K integral membr
25	195.5	15.2	218	1 A43522	CD9 antigen - bovi
26	189	14.7	226	1 JX0221	23K integral membr
27	188.5	14.7	218	1 A40181	tetraspan TSPAN-3
28	187.5	14.6	233	2 A59264	tetraspan TSPAN-2
29	184.5	14.4	222	2 A59263	

30	176.5	13.8	281	1 A47629	cell surface glyco
31	175.5	13.7	244	2 T13615	hypothetical prote
32	172.5	13.4	321	2 T45053	hypothetical prote
33	163.5	12.7	206	2 T25161	hypothetical prote
34	163	12.7	241	2 A59262	tetraspan TSPAN-1
35	162	12.6	242	2 T15361	hypothetical prote
36	150	11.7	233	2 T15620	hypothetical prote
37	149.5	11.7	223	2 T26763	hypothetical prote
38	142	11.1	210	2 T58391	sarcoma amplified
39	135.5	10.6	247	2 T28890	hypothetical prote
40	135	10.5	359	2 T18667	hypothetical prote
41	128.5	10.0	194	2 T25548	hypothetical prote
42	128	10.0	309	2 T34080	hypothetical prote
43	125	9.7	203	2 T22517	hypothetical prote
44	125	9.7	346	1 S10177	retinal degenerati
45	122	9.5	346	1 S03347	retinal degenerati

## ALIGNMENTS

### RESULT 1

A59261  
tetraspan TSPAN-5 - human  
C:Species: Homo sapiens (man)  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 02-Jun-2000  
C:Accession: A59261  
R: Todd, S.C.; Doctor, V.S.; Levy, S.  
Biochim. Biophys. Acta 1399, 101-104, 1998  
A:Title: Sequences and expression of six new members of the tetraspanin/TM4SF family.  
A:Reference number: A59258; MUID:98390278; PMID:9714763  
A:Accession: A59261  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-264 <TOD>  
A:Cross-references: GB:AF053455; NID:g2995864; PIDN:AAC69712.1; PID:g2995865  
C:Genetics:  
A:Gene: TSPAN-5  
C:Superfamily: CD9 antigen

Query Match 66.4%; Score 852; DB 2; Length 264;  
Best Local Similarity 77.8%; Pred. No. 6.5e-66;  
Matches 151; Conservative 16; Mismatches 21; Indels 6; Gaps 2;

QY	1	MPKGHQHQPPEVCGCKYF	FGFNIVFWVIGALFLAIGLWANGKGVLSNISALTDLGG	60
DB	1	MSGK--HYKGPVSCCIK	YIFGFNVIFWFLGITFLGIGLWANGKGVLSNISITDLGG	58
QY	61	LDPVWLFWVYGVVSGV	VLGFGAGCIGALRENTFLKFPFSVFLGLIFFLSLATGILAFVKDW	120
DB	59	FDPVWLFWVYGVVSGV	FLGFGAGCIGALRENTFLKFPFSVFLGLIFFLSLATGILAFVKDW	114
QY	121	IRDQLAPFNNNKA	VRDDIDLQNLIDFAOYWSCCGARGPNWNLNINFNCTDLPNSRE	180
DB	115	IKDQLYFFINNIR	AYRDDIDLQNLIDFTQYWCQCGAFGADWNLNINFNCTDLPNSRE	174
QY	181	RCGVPPSCCV	RDPA 194	
DB	175	RCGVPPSCCV	KDPA 188	

### RESULT 2

T24912  
hypothetical protein T14G10.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T24912  
R: Wild, A.  
submitted to the EMBL Data Library, January 1996  
A:Reference number: Z19954  
A:Accession: T24912  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA



QY 177 PSRRCGVFPSCC 189  
 Db 167 VYSEK-GPPKSCC 178

RESULT 6  
 139368  
 T-cell acute lymphoblastic leukemia associated antigen 1 - human  
 N:Alternate names: cell surface glycoprotein (clone A15); TALLA-1  
 C:Species: Homo sapiens (man)  
 C>Date: 23-Feb-1996 #sequence\_revision 23-Aug-1996 #text\_change 22-Jun-1999  
 C:Accession: 139368; 154784  
 R:Emm, N.; Kitoori, K.; Seto, M.; Ueda, R.; Saito, H.; Takahashi, T.  
 Immunogenetics 37, 193-198, 1993  
 A:Title: Isolation of a novel cDNA clone showing marked similarity to MB491/CD63 superfamily  
 A:Reference number: 139368; MUID:93131291; PMID:8420826  
 A:Accession: 139368  
 A:Molecule type: mRNA  
 A:Residues: 1-244 <RES>  
 A:Cross-references: GB:D10653; NID:g285900; PIDN:BAA01501.1; PID:g285901  
 A:Experimental source: immature T cell line HPB-ALL  
 R:Takagi, S.; Fujikawa, K.; Imai, T.; Fukuhara, N.; Fukudome, K.; Minegishi, M.; Tsuchiya, J.  
 Int. J. Cancer 61, 706-715, 1995  
 A:Title: Identification of a highly specific surface marker of T-cell acute lymphoblastic leukemia  
 A:Reference number: 154784; MUID:95286314; PMID:7769645  
 A:Accession: 154784  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-244 <RE>  
 A:Cross-references: GB:D29808; NID:g475005; PIDN:BAA06191.1; PID:g475006  
 C:Genetics:  
 A:Gene: GDB:MXS1; DXS1692E; A15; TALLA-1  
 A:Cross-references: GDB:202921  
 A:Map position: Xq11-Qx11  
 C:Superfamily: CD9 antigen  
 C:Keywords: glycoprotein; surface antigen; transmembrane protein  
 F:1-11/Domain: intracellular #status predicted <CY1>  
 F:12-35/Domain: transmembrane #status predicted <TM1>  
 F:36-51/Domain: extracellular #status predicted <EX1>  
 F:52-76/Domain: transmembrane #status predicted <TM2>  
 F:77-80/Domain: intracellular #status predicted <CY2>  
 F:81-102/Domain: transmembrane #status predicted <TM3>  
 F:103-207/Domain: extracellular #status predicted <EX2>  
 F:208-232/Domain: transmembrane #status predicted <TM4>  
 F:233-244/Domain: intracellular #status predicted <CY3>  
 F:49,150,153,172,183/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.7%; Score 240; DB 1; Length 244;  
 Best Local Similarity 28.2%; Pred. No. 2.3e-13;  
 Matches 50; Conservative 34; Mismatches 81; Indels 12; Gaps 2;

QY 16 CGKYFLFGNIVFWLGLAFLAIGLWANGKGVLSNLSALTDLGLDPVWLFWVVGVMVS 75  
 Db 9 CLKXTLLIYFVFWITGVILLAVGVWG---KLTGLTYSILAEINSTWAPVLTGTGTTIV 65

QY 76 VLGFAGCIGALRENTFLKFFSVFLGLIIFLELATGILAFVKDQIRDLNFFNNVKA 135  
 Db 66 VFLGFGCFATCRGSPWMLKLYANFLSLVFLAELVAGISGVFRHEIKDTFLRTYTDAMQT 125

QY 136 YRDDIDLQNLIDFAQYWSCGARGPNWNLNIFNCTDLNPSRCGVFPSCCVRD 192  
 Db 126 YNGNDEKRAVDHVQRSLSCGVQNTWNTSTPYF-----LEHGIPSCOMNE 173

RESULT 7  
 A59260  
 tetraspan TSPAN-6 - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 02-Jun-2000  
 C:Accession: A59260  
 R:Todd, S.C.; Doctor, V.S.; Levy, S.  
 Biochim. Biophys. Acta 1399, 101-104, 1998  
 A:Title: Sequences and expression of six new members of the tetraspanin/TM4SF family.

A:Reference number: A59258; MUID:98390278; PMID:9714763  
 A:Accession: A59260  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-245 <TOD>  
 A:Cross-references: GB:AF053454; NID:g2995862; PIDN:AAC69711.1; PID:g2995863  
 C:Genetics:  
 A:Gene: Tspan-6  
 C:Superfamily: CD9 antigen

Query Match 18.7%; Score 239.5; DB 2; Length 245;  
 Best Local Similarity 29.5%; Pred. No. 2.5e-13;  
 Matches 56; Conservative 32; Mismatches 89; Indels 13; Gaps 3;

QY 1 MPKGHQHFOEVEVCCGKYFLFGNIVFWLGLAFLAIGLWANGKGVLSNLSALTDLGG 60  
 Db 1 MASPSRLQTKPVITCLKSVLLIYTFIFWITGVILLAVGVWG---KVSLENYFSLNKA 57

QY 61 LDPVWLFWVVGVMVSVLGFAGCIGALRENTFLKFFSVFLGLIIFLELATGILAFVKDQ 120  
 Db 58 TNVPFVLGTGTVILLGTGFCPATCTSAWMLKLYANFLIIFLVELVAALVGVFVRHE 117

QY 121 IRDQLNFPINNKKAYRDDIDLQNLIDFAQYWSCGARGPNWNLNIFNCTDLNPSR 179  
 Db 118 IKNSFKSNYENALKEYNSTGDSYRSEAVDKIQSTLHCCGVTVNGDKGTYSET----- 171

QY 180 ERCGVFPSCC 189  
 Db 172 ---GPPKSCC 178

RESULT 8  
 A37243  
 hemopoietic cell surface glycoprotein CD53 - human  
 N:Alternate names: pan-leukocyte surface antigen CD53  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Dec-1991 #sequence\_revision 09-Aug-1996 #text\_change 22-Jun-1999  
 C:Accession: A37243; A45872  
 R:Amiot, M.  
 J. Immunol. 145, 4322-4325, 1990  
 A:Title: Identification and analysis of cDNA clones encoding CD53. A pan-leukocyte anti  
 A:Reference number: A37243; MUID:91079522; PMID:2258620  
 A:Accession: A37243  
 A:Molecule type: mRNA  
 A:Residues: 1-219 <AMI>  
 A:Cross-references: GB:M60871; NID:g180140; PIDN:AAA51951.1; PID:g180141  
 R:Angelisova, P.; Vlcek, C.; Stefanova, I.; Lipoldova, M.; Horejsi, V.  
 Immunogenetics 32, 281-285, 1990  
 A:Title: The human leukocyte surface antigen CD53 is a protein structurally similar to  
 A:Reference number: A45872; MUID:9105810; PMID:1700763  
 A:Accession: A45872  
 A:Molecule type: mRNA  
 A:Residues: 1-219 <ANG>  
 A:Cross-references: GB:M37033; NID:g180142; PIDN:AAA35663.1; PID:g180143  
 C:Genetics:  
 A:Gene: GDB:CD53; MOX44  
 A:Cross-references: GDB:127521; OMIM:151525  
 A:Map position: lp21-lp13.3  
 C:Superfamily: CD9 antigen  
 C:Keywords: glycoprotein; transmembrane protein  
 F:1-10/Domain: intracellular #status predicted <CY1>  
 F:11-36/Domain: transmembrane #status predicted <TM1>  
 F:37-54/Domain: extracellular #status predicted <EX1>  
 F:55-75/Domain: transmembrane #status predicted <TM2>  
 F:76-80/Domain: intracellular #status predicted <CY2>  
 F:81-106/Domain: transmembrane #status predicted <TM3>  
 F:107-181/Domain: extracellular #status predicted <EX2>  
 F:182-204/Domain: transmembrane #status predicted <TM4>  
 F:205-219/Domain: intracellular #status predicted <CY3>  
 F:129,148/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.4%; Score 235.5; DB 1; Length 219;  
 Best Local Similarity 31.0%; Pred. No. 5e-13;

```
Matches 54; Conservative 31; Mismatches 64; Indels 25; Gaps 5;
Qy 18 KYFLFGFNIVFWLGFALATGLW--ANGEKGVLSNLSALTDLGGLDPVWLVVVGVM 74
Db 10 KYVLFFNFWLFWGCCILGFIYLLIHNFGVLFNPLT-LGN-----VFVIGSII 63
Qy 75 SVLGFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDWIRDLQNLFFNNVK 134
Db 64 MVVAFLGCMGSIKENKCLMSFFILLILLAEVTLAILLFVYEKLNEYAKGLTDSIH 123
Qy 135 AYRDDIDLQNLIDFAQEWSCCGARGPNDNMLNIYFNCTDLNPSRRCGVFPSC 188
Db 124 RYHSDNSTKAAWDSIQSFLQCCGNGTSDWT-----SGPDASC 161

RESULT 9
A39574
leukocyte antigen OX-44 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Aug-1991 #sequence_revision 09-Aug-1996 #text_change 22-Jun-1999
C;Accession: A39574
R;Bellacosa, A.; Iazo, P.A.; Bear, S.E.; Tschlis, P.N.
Mol. Cell. Biol. 11, 2864-2872, 1991
A;Title: The rat leukocyte antigen MRC OX-44 is a member of a new family of cell surface
A;Reference number: A39574; MUID:91203909; PMID:2017181
A;Accession: A39574
A;Molecule type: mRNA
A;Residues: 1-219 <BEL>
A;Cross-references: GB:M57276; NID:G205897; PIDN:AAA41775.1; PID:G205898
C;Superfamily: CD9 antigen
F;1-10/Domain: intracellular #status predicted <CY1>
F;11-36/Domain: transmembrane #status predicted <CY1>
F;37-54/Domain: extracellular #status predicted <EX1>
F;55-73/Domain: transmembrane #status predicted <TM2>
F;74-79/Domain: intracellular #status predicted <CY2>
F;80-106/Domain: transmembrane #status predicted <TM3>
F;107-181/Domain: extracellular #status predicted <EX2>
F;182-204/Domain: transmembrane #status predicted <TM4>
F;205-219/Domain: intracellular #status predicted <CY3>
F;119,129,148/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.2%; Score 233.5; DB 1; Length 219;
Best Local Similarity 32.8%; Pred. No. 7.4e-13;
Matches 63; Conservative 24; Mismatches 70; Indels 35; Gaps 6;
Qy 18 KYFLFGFNIVFWLGFALATG--LWANGKGVLSNLSALTDLGGLDPVWLVVVGVM 74
Db 10 KYVLFFNFWLFWGCCILGFIHLLVQNTYIGILFNLPLT-LGN-----VLVIGSII 63
Qy 75 SVLGFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDWIRDLQNLFFNNVK 134
Db 64 MVVAFLGCMGSIKENKCLMSFFVLLILLAEVTLAILLFVYEKKINTLVAGINDSIQ 123
Qy 135 AYRDDIDLQNLIDFAQEWSCCGARGPNDNMLNIYFNCTDLNPSRRCGVFPSCVRDPA 194
Db 124 HYHSDNSTRAWDFIQSLOQCGVNGSSDW-----ISGPPSSC----- 161
Qy 195 MSSTPSPVAMMSG 206
Db 162 ----PSGADVQG 169

RESULT 10
A36056
tumor-associated antigen CO-029 - human
C;Species: Homo sapiens (man)
C;Date: 16-Nov-1990 #sequence_revision 09-Aug-1996 #text_change 20-Apr-2000
C;Accession: A36056
R;Szala, S.; Kasai, Y.; Steplewski, Z.; Rodeck, U.; Koprowski, H.; Linnenbach, A.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6833-6837, 1990
A;Title: Molecular cloning of cDNA for the human tumor-associated antigen CO-029 and id
A;Reference number: A36056; MUID:90370878; PMID:2395876
```

```
A;Accession: A36056
A;Molecule type: mRNA
A;Residues: 1-237 <SA>
A;Cross-references: GB:M35252; NID:G180925; PIDN:AAA35709.1; PID:G180926
C;Genetics:
A;Gene: GDB:TM4SF3
A;Cross-references: GDB:9113496; OMIM:600769
C;Superfamily: CD9 antigen
C;Keywords: glycoprotein; transmembrane protein
F;2-11/Domain: intracellular #status predicted <CY1>
F;12-33/Domain: transmembrane #status predicted <TM1>
F;34-52/Domain: extracellular #status predicted <EX1>
F;53-78/Domain: transmembrane #status predicted <TM2>
F;79-82/Domain: intracellular #status predicted <CY2>
F;83-107/Domain: transmembrane #status predicted <TM3>
F;108-205/Domain: extracellular #status predicted <EX2>
F;206-232/Domain: transmembrane #status predicted <TM4>
F;233-237/Domain: intracellular #status predicted <CY3>
F;37,118/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.0%; Score 231; DB 1; Length 237;
Best Local Similarity 33.1%; Pred. No. 1.3e-12;
Matches 56; Conservative 24; Mismatches 67; Indels 22; Gaps 4;
Qy 13 VGCGKYFLFGFNIVFWLGFALFIAIGLWA-----WCEKGVLSNLSALTDLGGLD 63
Db 4 VSACIKSYMTFFNFWLGCGLILLALAIWVWVNDSSQAFGSDVGSSVAVD----- 57
Qy 64 VMLFVVGVVSVLGFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDWIR 123
Db 58 --ILIAVGAIIMLGLGCCGAIKESCMILLFFIGLLILLQVATGILGAVFKSKSDR 115
Qy 124 QLNFFINNVKAY-----RDDIDLQNLIDFAQEWSCCG-ARGPNDNMLN 167
Db 116 IVNSTLYENTKLSATGESEKQFQEAIIVFQEEFKCCGLVNGAADWGN 164

RESULT 11
A46493
metastasis suppressor Kail - human
A;Alternate names: cell surface glycoprotein Kail; membrane protein R2, inducible; type
C;Species: Homo sapiens (man)
C;Date: 18-Jun-1993 #sequence_revision 09-Aug-1996 #text_change 22-Jun-1999
C;Accession: I38942; S16156; A46493
R;Dong, J.T.; Lamb, P.W.; Rinker-Schaeffer, C.W.; Vukanovic, J.; Ichikawa, T.; Isaacs,
Science 268, 884-886, 1995
A;Title: Kail, a metastasis suppressor gene for prostate cancer on human chromosome 11p
A;Reference number: I38942; MUID:95273964; PMID:7754374
A;Accession: I38942
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U20770; NID:G806805; PIDN:AAC50133.1; PID:G806806
R;Gaugitsch, H.W.; Hofer, E.; Huber, N.E.; Schnabl, E.; Baumrucker, T.
Eur. J. Immunol. 21, 377-383, 1991
A;Title: A new superfamily of lymphoid and melanoma cell proteins with extensive homolo
A;Reference number: S16156; MUID:91153380; PMID:1842498
A;Accession: S16156
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-267 <GAU>
A;Cross-references: EMBL:X51795; NID:G35832; PIDN:CAA37804.1; PID:G35833
A;Note: the authors translated the codon AGC for residue 50 as Thr
R;Imai, T.; Fukudome, K.; Takagi, S.; Nagira, M.; Furuse, M.; Fukuhara, N.; Nishimura, I.
J. Immunol. 149, 2879-2886, 1992
A;Title: C33 antigen recognized by monoclonal antibodies inhibitory to human T cell leu
D9, CD37, CD53, and CD63.
A;Reference number: A46493; MUID:93017900; PMID:1401919
A;Accession: A46493
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-239, 'MV', 242-267 <IMA>
A;Cross-references: GB:S48196; NID:G258294; PIDN:AAB23825.1; PID:G258295
```

A;Experimental source: T-cell line MOLT-4  
A;Note: sequence extracted from NCBI backbone (NCBIP:117149)  
C;Genetics:  
A;Gene: GDB:KAL1  
A;Cross-references: GDB:134216; OMIM:60623  
A;Map position: lip11.2-1lip11.2  
C;Superfamily: CD9 antigen  
C;Keywords: glycoprotein; transmembrane protein  
F;1-10/Domain: intracellular #status predicted <CY1>  
F;11-36/Domain: transmembrane #status predicted <TM1>  
F;37-57/Domain: extracellular #status predicted <EX1>  
F;58-78/Domain: transmembrane #status predicted <TM2>  
F;79-83/Domain: intracellular #status predicted <CY2>  
F;84-108/Domain: transmembrane #status predicted <TM3>  
F;109-227/Domain: extracellular #status predicted <EX2>  
F;228-252/Domain: transmembrane #status predicted <TM4>  
F;253-265/Domain: intracellular #status predicted <CY3>  
F;119,157,198/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 17.8%; Score 228; DB 1; Length 267;  
Best Local Similarity 30.0%; Pred. No. 2.7e-12;  
Matches 68; Conservative 37; Mismatches 88; Indels 34; Gaps 11;

QY 18 KYFLFGNIVFWLGLAFLAIGLWAGKEK-----GVLSNISALTDLGLDPLVFWVVGAV 73  
Db 10 KYFLFLFNLIFFILGAVILGFWILADKSFISVLQTSSTSSLRMG----AYVFIGVAV 65

QY 74 MSVLGAGGCGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWIRDLNFFINNV 133  
Db 66 TMLMGFLGCGAVNEVRCGLGLYFAFLLLILIAQVTAAGALFYFMNGKLKQEMGGIVTELI 125

QY 134 KAY---RDDIDLQNLIDFAQYWSCCGARGPNDWNLNIFNCTD----LNPSRRCGVFP 186  
Db 126 RDYNSRRED-SLQDWDVYQAVKCCG-----W-VSFYNWTDNAELMN--RPEVTYPC 174

QY 187 SCCVR---DPAMSTPVSAMSSNSWSRPRYTPKAVNASLRSGC 230  
Db 175 SCEVKGEEDNSLVKRGFCAPG-NRTQSGNHPEP-----MPVYQEGC 216

RESULT 12  
S43511  
CD63/ME491 antigen homolog - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 13-Jan-1995 #sequence\_revision 09-Aug-1996 #text\_change 16-Jun-2000  
C;Accession: S43511  
R;Miyamoto, H.; Homma, M.; Hotta, H.  
Biochem. Biophys. Acta 1217, 312-316, 1994  
A;Title: Molecular cloning of the murine homologue of CD63/ME491 and detection of its st  
A;Reference number: S43511; MUID:94198294; PMID:8148377  
A;Accession: S43511  
A;Molecule type: mRNA  
A;Residues: 1-238 <MY>  
A;Cross-references: EMBL:D16432; NID:9484052; PIDN:BAA03904.1; PID:9976238  
C;Superfamily: CD9 antigen  
C;Keywords: glycoprotein; lysosome; surface antigen; transmembrane protein  
F;1-17/Domain: intracellular #status predicted <CY1>  
F;18-35/Domain: transmembrane #status predicted <TM1>  
F;36-51/Domain: extracellular #status predicted <EX1>  
F;52-76/Domain: transmembrane #status predicted <TM2>  
F;77-80/Domain: intracellular #status predicted <CY2>  
F;81-103/Domain: transmembrane #status predicted <TM3>  
F;104-202/Domain: extracellular #status predicted <EX2>  
F;203-228/Domain: transmembrane #status predicted <TM4>  
F;229-238/Domain: intracellular #status predicted <CY3>  
F;116,130,150,172/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 17.1%; Score 220; DB 1; Length 238;  
Best Local Similarity 31.2%; Pred. No. 1.2e-11;  
Matches 55; Conservative 31; Mismatches 76; Indels 14; Gaps 5;

QY 16 CKYKFLFGNIVFWLGLAFLAIGLWAGKEKGVLSN-ISALTDLGLDPLVFWVVGAV 74

Db 9 CVKFLLYVLLAFACACAVGLIAGV---AVQVVLKQAI THETTAGSLLPV-VIIAVGAFL 64

QY 75 SVLGFAAGCAGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWIRDLNFFINNVK 134  
Db 65 FLVAFVCCGACKENYCLMITFAIFLSLIMLVEVAIAIGYVFRDQVKSEFNKSFQOQMQ 124

QY 135 AYRDDIDLQNLIDFAQYWSCCGARGPNDWNLNIFNCTDNLNPSRRCGVFPFSCCV 190  
Db 125 NYLKDKNTATILDKLQKNNCCGASNYDWE-NI-----PGMAKDRVPDSCCI 171

RESULT 13  
I49561  
C33/R2/IA4 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Aug-1999  
C;Accession: I49561  
R;Nagira, M.; Imai, T.; Ishikawa, I.; Uwabe, K.I.; Yoshie, O.  
Cell. Immunol. 157, 144-157, 1994  
A;Title: Mouse homologue of C33 antigen (CD82), a member of the transmembrane 4 superf  
A;Reference number: I49561; MUID:94313678; PMID:8039242  
A;Accession: I49561  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-266 <RES>  
A;Cross-references: GB:D14883; NID:9984036; PIDN:BAA03602.1; PID:9498368  
C;Superfamily: CD9 antigen

Query Match 16.9%; Score 217; DB 2; Length 266;  
Best Local Similarity 31.5%; Pred. No. 2.4e-11;  
Matches 51; Conservative 32; Mismatches 67; Indels 12; Gaps 4;

QY 14 GC--CKYFLFGNIVFWLGLAFLAIGLWAGKEK-----GVLSNISALTDLGLDPLVWLF 67  
Db 4 GCYKVTKYFLFLFNLIFFILGAVILGFWILADKNSFISVLQTSSTSSQLVG----AYVF 59

QY 68 VVVGWMSVLGFGAGCGALRENTFLKFFSVFLGLIFFLELATGILAFVKDWIRDLNLF 127  
Db 60 IGVGALTIVMGFLGCGAVNEVRCGLGLYFVFLLLILIAQVTVGVLFYFNADKLKEMGN 119

QY 128 FINNVKAYRDDI--DLQNLIDFAQYWSCCGARGPNDWNLN 167  
Db 120 TVMDIIRNTANATSSREAWDYVQAVKCCGWVSHYNWTEN 161

RESULT 14  
I38016  
melanoma-associated antigen CD63 (validated) - human  
N;Alternate names: antigen ME491; lysosomal membrane glycoprotein CD63; ME491/CD63 ant  
C;Species: Homo sapiens (man)  
C;Date: 17-May-1996 #sequence\_revision 09-Aug-1996 #text\_change 08-Dec-2000  
C;Accession: I38016; S01418; A39514; B35826; A61177; A61173; A56782  
R;Hotta, H.; Miyamoto, H.; Hara, I.; Takahashi, N.; Homma, M.  
Biochem. Biophys. Res. Commun. 185, 436-442, 1992  
A;Title: Genomic structure of the ME491/CD63 antigen gene and functional analysis of t  
A;Reference number: I38016; MUID:92287132; PMID:1599482  
A;Accession: I38016  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-238 <RES>  
A;Cross-references: EMBL:X62654; NID:9430755; PIDN:CAA44519.1; PID:9430756  
R;Hotta, H.; Ross, A.H.; Huebner, K.; Isobe, M.; Wendeborn, S.; Chao, M.V.; Ricciardi,  
Cancer Res. 48, 2955-2962, 1988  
A;Title: Molecular cloning and characterization of an antigen associated with early st  
A;Reference number: S01418; MUID:88210273; PMID:3365686  
A;Accession: S01418  
A;Molecule type: mRNA  
A;Residues: 1-238 <HOT>  
A;Cross-references: EMBL:X07982; NID:934526; PIDN:CRA30792.1; PID:g34527  
R;Metzelaar, M.J.; Wijngaard, P.L.J.; Peters, P.J.; Sixma, J.J.; Nieuwenhuis, H.K.; Cl  
J. Biol. Chem. 266, 3239-3245, 1991  
A;Title: CD63 antigen. A novel lysosomal membrane glycoprotein, cloned by a screening  
A;Reference number: A39514; MUID:91131632; PMID:1993697

A;Accession: A39514  
A;Molecule type: mRNA  
A;Residues: 1-238 <MET>  
A;Cross-references: GB:M58485  
R;Rapp, G.; Freudenstein, J.; Klaudiny, J.; Mucha, J.; Wenpe, F.; Zimmer, M.; Scheit, K.  
DNA Cell Biol. 9, 479-485, 1990  
A;Title: Characterization of three abundant mRNAs from human ovarian granulosa cells.  
A;Reference number: A35826; MUID:91025550; PMID:2171551  
A;Accession: B35826  
A;Molecule type: mRNA  
A;Residues: 1-238 <RAP>  
A;Cross-references: GB:M59907; NID:9189383; PIDN:AAA63235.1; PID:9189384  
A;Note: the authors do not translate the codons for residues 205 through 224  
R;Azorsa, D.O.; Hyman, J.A.; Hildreth, J.E.K.  
Blood 78, 280-284, 1991  
A;Title: CD63/Pltgp40: a platelet activation antigen identical to the stage-specific, me  
A;Reference number: A61177; MUID:91300080; PMID:2070066  
A;Accession: A61177  
A;Status: nucleic acid sequence not shown; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 2-68, 'P', 70-238 <AZO>  
R;Hildreth, J.E.K.; Derr, D.; Azorsa, D.O.  
Blood 77, 121-132, 1991  
A;Title: Characterization of a novel self-associating Mr 40,000 platelet glycoprotein.  
A;Reference number: A61173; MUID:91084576; PMID:1984792  
A;Accession: A61173  
A;Molecule type: protein  
A;Residues: 2-8, 'X', 10-16, 'XX', 19-21 <HIL>  
R;Wang, M.X.; Earley Jr., J.J.; Shields, J.A.; Donoso, L.A.  
Arch. Ophthalmol. 110, 399-404, 1992  
A;Title: An ocular melanoma-associated antigen. Molecular characterization.  
A;Reference number: A56782; MUID:92181348; PMID:1339263  
A;Accession: A56782  
A;Molecule type: mRNA  
A;Residues: 1-238 <WAN>  
A;Cross-references: GB:893788; NID:9246538; PIDN:AA21617.1; PID:9246539  
A;Experimental source: uveal melanoma  
A;Note: sequence extracted from NCBI backbone (NCBIN:93788, NCBI:P:93790)  
C;Genetics:  
A;Gene: GDB:CD63; MLA1  
A;Cross-references: GDB:120186; OMIM:155740  
A;Map position: 12q12-12q13  
A;Introns: 22/3; 85/3; 110/3; 142/3; 189/3; 217/3  
C;Superfamily: CD9 antigen  
C;Keywords: glycoprotein; lysosome; surface antigen; transmembrane protein  
F;2-238/Product: melanoma-associated antigen M2491 #status experimental <MAT>  
F;2-11/Domain: intracellular #status predicted <CY1>  
F;12-35/Domain: transmembrane #status predicted <TM1>  
F;36-51/Domain: extracellular #status predicted <EX1>  
F;52-76/Domain: transmembrane #status predicted <TM2>  
F;77-80/Domain: intracellular #status predicted <CY2>  
F;81-103/Domain: transmembrane #status predicted <TM3>  
F;104-202/Domain: extracellular #status predicted <EX2>  
F;203-228/Domain: transmembrane #status predicted <TM4>  
F;229-238/Domain: intracellular #status predicted <CY3>  
F;130,150,172/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 16.7%; Score 214; DB 1; Length 238;  
Best Local Similarity 28.1%; Pred. No. 3.8e-11;  
Matches 61; Conservative 37; Mismatches 83; Indels 36; Gaps 8;  
QY 16 CGKYFLFGFNIVFWLGFALFLAIGLWAWGKGVLSN--ISALTDLGLDLPVWLVFVVG 73  
Db 9 CVKFLLYVLLAFACACAVGLIAGV---GAQLVLSQTIIGATP-GSLLPV-VIIAGV 63  
QY 74 MSVLGAGCIGALRENTFLKFRSVFLGLIFFLELATGILAFVKDWIRDLQNFNNV 133  
Db 64 LFLVAFVGGCCGACKENYCLMITEFAIFLSLIMLVEVAAGIAGYVFRDKVMSEFNFRQM 123  
QY 134 KAYRDDIDLQNLIDFAQYWSCCGARGPNWNLNIFNCTDLNPSRRCGVPPSCV 193  
Db 124 ENPKNNHTASILDRWQADPKCCGAANYTDW-----EKTPSMKRNRPVDSCCI--- 171

QY 194 AMSSTPSVAMWSSNWSWSRAPYTPKAVWASLRSGC 230  
Db 172 -----NVTGCGGIN-----FNEKAIH---KEGC 191  
RESULT 15  
JC2297  
CD63 antigen - rabbit  
C;Species: Oryctolagus cuniculus (domestic rabbit)  
C;Date: 28-Oct-1994 #sequence\_revision 09-Aug-1996 #text\_change 16-Jun-2000  
C;Accession: JC2297  
R;Schma, Y.; Suzuki, T.; Sasano, H.; Nagura, H.; Nose, M.; Yamamoto, T.  
Cell Struct. Funct. 19, 219-225, 1994  
A;Title: Increased mRNA for CD63 antigen in atherosclerotic lesions of Watanabe herit  
A;Reference number: JC2297; MUID:95120837; PMID:7820873  
A;Accession: JC2297  
A;Molecule type: mRNA  
A;Residues: 1-238 <SOH>  
A;Cross-references: DDBJ:D21264; NID:9684973; PIDN:BA04804.1; PID:9684974  
A;Experimental source: aorta  
C;Superfamily: CD9 antigen  
C;Keywords: glycoprotein; lysosome; surface antigen; transmembrane protein  
F;1-11/Domain: intracellular #status predicted <CY1>  
F;12-35/Domain: transmembrane #status predicted <TM1>  
F;36-51/Domain: extracellular #status predicted <EX1>  
F;52-76/Domain: transmembrane #status predicted <TM2>  
F;77-80/Domain: intracellular #status predicted <CY2>  
F;81-103/Domain: transmembrane #status predicted <TM3>  
F;104-202/Domain: extracellular #status predicted <EX2>  
F;203-228/Domain: transmembrane #status predicted <TM4>  
F;229-238/Domain: intracellular #status predicted <CY3>  
F;125,130,150,172/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 16.7%; Score 214; DB 1; Length 238;  
Best Local Similarity 31.2%; Pred. No. 3.8e-11;  
Matches 55; Conservative 32; Mismatches 75; Indels 14; Gaps 5;  
QY 16 CGKYFLFGFNIVFWLGFALFLAIGLWAWGKGVLS-NISALTDLGLDLPVWLVFVVG 74  
Db 9 CVKFLLYVLLAFACACAVGLIAGV---GAQLVLSQTIIGATP-GSLLPV-VIIAGV 64  
QY 75 SVLGFAGCIGALRENTFLKFRSVFLGLIFFLELATGILAFVKDWIRDLQNFNNV 134  
Db 65 FLVAFVGGCCGCTCKENYCLMITEFAIFLSLIMLVEVAAGIAGYVFRDKVMSEFN 124  
QY 135 AYRDDIDLQNLIDFAQYWSCCGARGPNWNLNIFNCTDLNPSRRCGVPPSCV 190  
Db 125 NYSTDNQTALILDRWQADPKCCGAANYTDW-----ATPGWTRDR--VPDSCV 171  
Search completed: November 21, 2003, 13:45:26  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: November 21, 2003, 13:34:37 ; Search time 40 Seconds  
(without alignment)  
924.582 Million cell updates/sec

Title: US-09-972-970-4

Perfect score: 1283

Sequence: 1 MPCKHQHFQEPVCGCGKYF.....RAPPYTKAWASLRSGCRTT 233

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1283	100.0	233	AAE13218	Human gene 1 encod
2	1066	83.1	270	AAAB49504	Clone HNTMH27. Ho
3	1066	83.1	270	AAAB47950	Human 23228. Homo
4	1066	83.1	329	AAAB18969	Amino acid sequenc
5	1055	82.2	329	AAAB90429	Human polypeptide
6	972.5	75.8	330	ABP69599	Human polypeptide
7	885	69.0	271	AAAB53438	Human colon cancer
8	791	61.7	268	AAAB49505	Clone HE8EJ16 #1.
9	699	54.5	193	AAAY42381	Amino acid sequenc

10	699	54.5	193	22	AAU39073	Human secreted pro
11	699	54.5	193	23	ABB55782	Human polypeptide
12	640.5	49.9	512	22	ABP98695	Rat gamma-hydroxyb
13	638.5	49.8	270	21	AAI76266	Human secreted pro
14	638.5	49.8	270	22	AAAB7034	Human secreted pro
15	638.5	49.8	270	23	AAE18535	Human tetraspan pr
16	637.5	49.7	270	22	AAAB7134	Human secreted pro
17	635.5	49.5	270	22	AAAB7135	Human secreted pro
18	635.5	49.5	292	22	ABG27829	Novel human diagno
19	634.5	49.5	270	22	AAAB7136	Human secreted pro
20	634.5	49.5	270	22	AAAB7137	Human secreted pro
21	625.5	48.8	270	21	AAI76133	Human secreted pro
22	619.5	48.3	270	21	AAAB23036	Human tetraspanin-
23	614.5	47.9	253	22	AAAB7039	Human TANGO 339 pe
24	556	43.3	200	22	ABG15480	Novel human diagno
25	515.5	40.2	228	22	AAAB7035	Human mature TANGO
26	482.5	37.6	122	21	AAI83082	F-box protein FBP-
27	482.5	37.6	122	23	AAO22468	Human F-box protei
28	482.5	37.6	269	22	ABB64234	Drosophila melanog
29	471.5	36.7	283	22	AAU27652	Human protein AFP2
30	471.5	36.7	283	23	ABP69619	Human polypeptide
31	453.5	35.3	193	22	AAAB7037	Human TANGO 339 tr
32	437.5	34.1	368	23	ABB89649	Human polypeptide
33	429	33.4	120	22	AAI75522	Human colon cancer
34	425	33.1	213	21	AAAB12120	Hydrophobic domain
35	421.5	32.9	291	22	ABB60781	Drosophila melanog
36	412.5	32.2	244	22	AAE06071	Human gene 31 enco
37	412.5	32.2	244	23	ABG33893	Human secreted pro
38	412.5	32.2	245	21	AAI87094	Human secreted pro
39	411.5	32.1	209	21	AAAB23047	Human tetraspanin-
40	411.5	32.1	304	22	ABB65216	Drosophila melanog
41	383	29.9	111	22	AAAB49510	Clone HE8EL16 #2.
42	373	29.1	248	22	ABG15481	Novel human diagno
43	331	25.8	104	21	AAI64846	Human 5' EST relat
44	324	25.3	294	19	AAI70319	Secreted protein B
45	324	25.3	294	21	AAI99374	Human PRO1311 (UNQ

## ALIGNMENTS

### RESULT 1

AF\_13218

ID AAE13218 standard; Protein; 233 AA.

XX AAE13218;

AC AAE13218;

DT 12-FEB-2002 (first entry)

XX Human gene 1 encoded TM4SF receptor protein HOF0B55, SEQ ID NO:4.

Human; 4-transmembrane superfamily; TM4SF receptor; breast cancer; ovarian cancer; immune disorder; Addison's disease; wound healing; gene therapy; autoimmune haemolytic anaemia; autoimmune thyroiditis; diabetes mellitus; allergy; Crohn's disease; multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disorder; myocardial ischaemia; neurological disease; antifungal; antiviral; antibacterial; cerebral anoxia; epilepsy; infectious disease; antiparasitic; cancer.

OS Homo sapiens.

XX

XX

Key Location/Qualifiers

Region 158..164

FT /note= "Immunogenic epitope"

Region 174..182

FT /note= "Immunogenic epitope"

Region 212..218

FT /note= "Immunogenic epitope"

Region 228..233

FT /note= "Immunogenic epitope"

XX WO200177173-A1.



XX 18-OCT-2001.  
 PD  
 XX 05-APR-2001; 2001WO-US11130.  
 XX  
 PF  
 XX 10-APR-2000; 2000US-195336P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Shi Y, Ruben SM;  
 PI  
 XX WPI; 2002-017447/02.  
 XX  
 DR N-PSDB; AAD21883.  
 DR  
 XX Novel isolated protein, a member of 4-transmembrane superfamily of  
 PT receptor polypeptides, useful in the prevention, treatment and  
 PT diagnosis of cancer, immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 XX  
 XX Claim 11; Page 267-268; 271pp; English.  
 PS  
 XX The invention relates to human 4-transmembrane superfamily (TM4SF)  
 CC receptor polypeptides and polynucleotides. Sequences of the invention  
 CC are useful for preventing, treating, ameliorating or diagnosing a  
 CC pathological condition or a susceptibility to a pathological condition.  
 CC TM4SF polypeptides are useful for screening molecules which modify  
 CC their activity. TM4SF nucleic acids, protein, antibodies, agonists and  
 CC antagonists are useful in the diagnosis, treatment and prevention of  
 CC cancer, particularly breast and ovarian cancer, and other cancers of  
 CC the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,  
 CC liver, lung, or urogenital; immune disorders such as Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; cardiovascular disorders such as  
 CC myocardial ischaemias; wound healing; neurological diseases such as  
 CC cerebral anoxia and epilepsy; and infectious diseases such as viral,  
 CC bacterial, fungal and parasitic infections. TM4SF polynucleotides are  
 CC also useful in gene therapy. The present sequence is human TM4SF  
 CC receptor protein.  
 XX  
 XX Sequence 233 AA;  
 SQ  
 Query Match 100.0%; Score 1283; DB 23; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-128;  
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MPKQHOFQPEPEVGGCGKYLFGFNIVFWLGLFLAIGLWAWGKGVLSNLSALTDLGG 60  
 DB 1 MPKQHOFQPEPEVGGCGKYLFGFNIVFWLGLFLAIGLWAWGKGVLSNLSALTDLGG 60  
 QY 61 LDPVWLVVVVGGVMSVLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
 DB 61 LDPVWLVVVVGGVMSVLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
 QY 121 IRDLNFFNNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDWNINIFNCTDLNPSRE 180  
 DB 121 IRDLNFFNNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDWNINIFNCTDLNPSRE 180  
 QY 181 RCGVPSCCVRDPMSTSPVAMSGSNWSRSPYTPKAVWASLRSGRTT 233  
 DB 181 RCGVPSCCVRDPMSTSPVAMSGSNWSRSPYTPKAVWASLRSGRTT 233  
 RESULT 2  
 AAB49504  
 ID AAB49504 standard; Protein; 270 AA.  
 XX  
 AC AAB49504;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Clone HNTMH27.  
 XX

KW Human: 4 transmembrane superfamily receptor protein;  
 KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
 KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
 KW bacterial infection; viral; fungal.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200070076-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 18-MAY-2000; 2000WO-US13504.  
 XX  
 PR 19-MAY-1999; 99US-0135122.  
 PR 03-JUN-1999; 99US-0137797.  
 PR 11-JUN-1999; 99US-0138573.  
 PR 18-AUG-1999; 99US-0149447.  
 PR 28-JAN-2000; 2000US-0178770.  
 XY  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;  
 PI Rosen CA;  
 XX  
 DR WPI; 2001-007502/01.  
 DR N-PSDB; AAC90014.  
 XX  
 XX Isolated nucleic acid molecule encoding human soluble 4 transmembrane  
 PT superfamily receptor protein, useful for diagnosing, treating and/or  
 PT preventing disorders e.g. Alzheimer's, cancer and arrhythmia -  
 XX  
 XX Claim 11; Pages 280-281; 297pp; English.  
 CC The present invention relates to isolated nucleic acids and proteins  
 CC encoding human soluble 4 transmembrane superfamily receptor protein (see  
 CC AAC90012-C90023 and AAB49502-B49513). The present sequence is one such  
 CC protein. The present protein can be used to screen for binding partners  
 CC and molecules which modify its activity. Antibodies specific for the  
 CC present protein can be used to treat and/or prevent diseases associated  
 CC with aberrant expression or activity of the present protein e.g.  
 CC endocrine disorders e.g. Addison's disease, (cardio)vascular diseases  
 CC e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural  
 CC disorders e.g. Alzheimer's and Parkinson's disease, reproductive  
 CC disorders, skin disorders e.g. psoriasis, renal system disorders e.g.  
 CC nephritis, (auto)immune system disorders e.g. graft vs. host disease,  
 CC hyperproliferative disorders e.g. neoplasms of the pancreas, ocular  
 CC disorders e.g. glaucoma and infections caused by bacteria, viruses and  
 CC fungi.  
 XX  
 SQ Sequence 270 AA;  
 Query Match 83.1%; Score 1066; DB 22; Length 270;  
 Best Local Similarity 99.5%; Pred. No. 1.2e-104;  
 Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MPKQHOFQPEPEVGGCGKYLFGFNIVFWLGLFLAIGLWAWGKGVLSNLSALTDLGG 60  
 DB 1 MPKQHOFQPEPEVGGCGKYLFGFNIVFWLGLFLAIGLWAWGKGVLSNLSALTDLGG 60  
 QY 61 LDPVWLVVVVGGVMSVLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
 DB 61 LDPVWLVVVVGGVMSVLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
 QY 121 IRDLNFFNNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDWNINIFNCTDLNPSRE 180  
 DB 121 IRDLNFFNNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDWNINIFNCTDLNPSRE 180  
 QY 181 RCGVPSCCVRDPA 194  
 DB 181 RCGVPSCCVRDPA 194  
 RESULT 3

AAB47950  
ID AAB47950 standard; Protein; 270 AA.  
XX  
AC AAB47950;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Human 23228.  
XX  
KW Tetraspanin; 23228; cell surface protein; transmembrane domain;  
KW extracellular growth factor; HB-EGF; TGF-alpha; amphiregulin;  
KW diabetes mellitus; arthritis; multiple sclerosis; encephalomyelitis;  
KW dermatitis; Crohn's disease; and asthma; cancer; metastasis;  
KW viral infection; cellular signalling activity; cell proliferation;  
KW cell motility; CD81; B-Cell antigen receptor.  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Domain 1..18 /label= N-terminal\_domain  
FT FT 18..263 /label= Tetraspanin\_domain  
FT FT 19..43 /label= TM1  
FT FT 44..63 /note= "Extracellular loop"  
FT FT 47..52 /note= "N-myristoylated"  
FT FT 51..54 /note= "N-glycosylated"  
FT FT 64..86 /label= TM2  
FT FT 71..76 /note= "N-myristoylated"  
FT FT 81..86 /note= "N-myristoylated"  
FT FT 87..94 /note= "N-myristoylated"  
FT FT 95..117 /note= "Intracellular loop"  
FT FT /label= TM3  
FT FT 118..234 /note= "Extracellular loop"  
FT FT 171..174 /note= "N-glycosylated"  
FT FT 183..188 /note= "N-myristoylated"  
FT FT 235..256 /label= TM4  
FT FT 240..245 /note= "N-myristoylated"  
FT FT 252..257 /note= "N-myristoylated"  
FT FT 257..270 /label= C-terminal\_domain  
XX  
PN W0200216603-A2.  
XX  
PD 28-FEB-2002.  
XX  
XX 21-AUG-2001; 2001WO-US41811.  
XX  
XX 21-AUG-2000; 2000US-226612P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Leiby KR;  
XX  
XX WPI; 2002-315466/35.  
DR N-PSDB; AAI72633.  
XX  
XX New tetraspanin 23228 polypeptide useful in screening assays,  
XX predictive medicine and as a prophylactic or therapeutic agent, e.g.,  
PT

PT for hematopoietic and immune diseases such as diabetes or multiple  
PT sclerosis -  
XX  
PS Claim 1; Page 113-114; 123pp; English.  
XX  
CC This sequence shows tetraspanin 23228. This protein is a cell surface  
CC protein having four transmembrane domains (TM1-TM4). TM1, TM3 and TM4  
CC have a single polar amino acid located within them, which may interact  
CC with each other and contribute to domain stability. The cytoplasmic N-  
CC and C-terminal domains and the intracellular loop between TM2 and TM3  
CC are less than 30 amino acids in length. The second extracellular domain  
CC between TM3 and TM4 contains conserved Cys residues and may function  
CC to bind extracellular growth factors, such as HB-EGF, TGF-alpha and  
CC amphiregulin. The 23228 protein is useful for diagnosing and treating  
CC 23228-mediated disorders, e.g., haematopoietic and/or immune disorders  
CC such as diabetes mellitus, arthritis, multiple sclerosis,  
CC encephalomyelitis, dermatitis, Crohn's disease and asthma; cell  
CC proliferation and differentiation disorders, e.g., cancers or  
CC metastasis; and/or viral infections. 23228 polypeptide can control  
CC cellular signalling activity, bind to an extracellular growth factor,  
CC for example, amphiregulin, regulate cell proliferation, bind to a cell  
CC surface protein, to recruit intracellular kinases, to regulate cell  
CC motility, bind to another tetraspanin such as CD81, to associate with a  
CC B-Cell antigen receptor and the ability to modulate the association  
CC with a virus.  
XX  
SQ Sequence 270 AA;  
Query Match 83.1%; Score 1066; DB 23; Length 270;  
Best Local Similarity 99.5%; Pred. No. 1.2e-104;  
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MPKGQHOFQEPVEVGGCCGKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNLSALTDLGG 60  
DB 1 MPKGQHOFQEPVEVGGCCGKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNLSALTDLGG 60  
QY 61 LDPVWLFWVVGWMSVLGFGAGCIGALRENTFLKFFSVFLGLFFLELATGILAFVKDW 120  
DB 61 LDPVWLFWVVGWMSVLGFGAGCIGALRENTFLKFFSVFLGLFFLELATGILAFVKDW 120  
QY 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQYVWSCGARGPNDNLIYFNCTDLNPSRE 180  
DB 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQYVWSCGARGPNDNLIYFNCTDLNPSRE 180  
QY 181 RGVPPSCCVRDPA 194  
DB 181 RGVPPSCCVRDPA 194  
RESULT 4  
AAB18969  
ID AAB18969 standard; Protein; 329 AA.  
XX  
AC AAB18969;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Amino acid sequence of a human transmembrane protein.  
XX  
KW Human; transmembrane protein; cell proliferation disorder; myeloma;  
KW reproductive disorder; smooth muscle disorder; neurological disorder;  
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;  
KW Alzheimer's disease; Tourette's disorder.  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Modified-site 51 /note= "potential glycosylation site"  
FT FT 171 /note= "potential glycosylation site"  
FT FT 269 /note= "potential glycosylation site"

FT Modified-site /note= "potential glycosylation site"  
285  
FT Modified-site /note= "potential phosphorylation site"  
325  
FT Modified-site /note= "potential phosphorylation site"  
XX PN WO2000056891-A2.  
XX PD 28-SEP-2000.  
XX PF 22-MAR-2000; 2000WO-US07817.  
XX PR 22-MAR-1999; 99US-0125537.  
XX PR 16-JUN-1999; 99US-0139565.  
XX PA (INCY-) INCYTE PHARM INC.  
XX PI Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O, Baughn MR;  
PI Lu DAM, Azimzai Y, Yang J;  
XX DR WPI; 2000-579485/54.  
XX DR N-PSDB; AAA96482.  
XX PT New human transmembrane proteins are used to treat a disease or  
PT condition associated with decreased expression of functional HTPM e.g.  
PT Tourette's disorder, angina and leukaemia -  
XX PS Claim 1; Page 91; 130pp; English.  
XX CC The present sequence represents a human transmembrane proteins (HTMP).  
CC Agonists and antagonists of the protein are used to treat a disease  
CC or condition associated with overexpression of the protein. Diseases  
CC and conditions which can be treated include cell proliferative,  
CC immunological, reproductive, smooth muscle and neurological disorders  
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency  
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,  
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The  
CC polynucleotides may be used to detect and quantify gene expression in  
CC biopsied tissues where protein expression may be correlated with disease  
CC e.g. to determine absence, presence or excess expression of HTPM or to  
CC monitor regulation of HTPM expression during therapeutic intervention.  
XX SQ Sequence 329 AA;  
Query Match 83.1%; Score 1066; DB 21; Length 329;  
Best Local Similarity 99.5%; Pred. No. 1.6e-104;  
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MPKQHQHFQPEVGGCCGKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNLSALTDLGG 60  
DB 1 MPKQHQHFQPEVGGCCGKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNLSALTDLGG 60  
QY 61 LDPVWLFWVVGVMVSLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
DB 61 LDPVWLFWVVGVMVSLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
QY 121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDNWNINIFNCTDLNPSRE 180  
DB 121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDNWNINIFNCTDLNPSRE 180  
QY 181 RCGVPFSCCVRDPA 194  
DB 181 RCGVPFSCCVRDPA 194  
RESULT 5  
ABB90429  
ID ABB90429 standard; Protein; 329 AA.  
XX AC ABB90429;  
XX AC ABB90429;  
DT 24-MAY-2002 (first entry)  
XX

DE Human polypeptide SEQ ID NO 2805.  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX OS Homo sapiens.  
XX PN WO2001190304-A2.  
XX PD 29-NOV-2001.  
XX PF 18-MAY-2001; 2001WO-US16450.  
XX PR 19-MAY-2000; 2000US-205515P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Birse CE, Rosen CA;  
XX DR WPI; 2002-122018/16.  
XX DR N-PSDB; ABL90838.  
XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders -  
XX PS Claim 11; SEQ ID NO 2805; 2081pp + Sequence Listing; English.  
XX CC The invention relates to novel genes (ABL90449-ABL90853) and proteins  
CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or uterine;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 329 AA;  
Query Match 82.2%; Score 1055; DB 23; Length 329;  
Best Local Similarity 98.5%; Pred. No. 2.3e-103;  
Matches 191; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MPKQHQHFQPEVGGCCGKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNLSALTDLGG 60  
DB 1 MPKQHQHFQPEVGGCCGKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNLSALTDLGG 60  
QY 61 LDPVWLFWVVGVMVSLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
DB 61 LDPVWLFWVVGVMVSLGFGAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVKDW 120  
QY 121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDNWNINIFNCTDLNPSRE 180  
DB 121 IRDQLNFFINNNVKAYRDDIDLQNLIDFAQEWSCCGARGPNDNWNINIFNCTDLNPSRE 180  
QY 181 RCGVPFSCCVRDPA 194  
DB 181 RCGVPFSCCVRDPA 194

RESULT 6  
ABP69599  
ID ID ABP69599 standard; Protein; 330 AA.  
XX AC  
XX ABP69599;  
XX 20-JAN-2003 (first entry)  
XX Human polypeptide SEQ ID NO 1646.  
XX  
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;  
XX cell-proliferative disorder; neurodegenerative disease; bacterial;  
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;  
XX multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
XX arthritis; cytostatic; immunomodulator; neutropenic; neuroprotective;  
XX antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
XX haemostatic; vulnary; fungicide; antibacterial; virucide; protozoicide;  
XX antiarthritic.  
XX  
XX Homo sapiens.  
XX  
XX WO200270539-A2.  
XX  
XX 12-SEP-2002.  
XX  
XX 05-MAR-2002; 2002WO-US05095.  
XX  
XX 05-MAR-2001; 2001US-0799451.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren P;  
XX Yue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
XX Wehrman T, Wang J, Wang D, Drmanac RT;  
XX  
XX WPI; 2002-759812/82.  
XX  
XX N-PSDB; AB211816.  
XX  
XX New polynucleotides comprising sequences assembled from expressed  
XX sequence tags (ESTs), useful for treating cell-proliferative, or  
XX neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
XX platelet or coagulation disorders -  
XX  
XX Claim 9; SEQ ID NO 1646; 1012pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (I) comprising a  
XX nucleotide sequence selected from any of 948 sequences  
XX (AB211119-AB212066) or their mature protein coding portion, active domain  
XX coding protein or complementary sequences. The polynucleotides are useful  
XX for identifying expressed genes or for physical mapping of human genome.  
XX The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
XX weight markers, as a food supplement, for generating antibodies, in  
XX medical imaging, screening and diagnostic assays and for treating  
XX cell-proliferative disorders (cancer), neurodegenerative diseases  
XX (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
XX sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
XX disorders, platelet or coagulation disorders, wound, burns, incision,  
XX ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
XX parasitic), arthritis, etc.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 330 AA;  
XX  
XX Query Match 75.8%; Score 972.5; DB 23; Length 330;  
XX Best Local Similarity 91.9%; Pred. NO. 1.3e-94;  
XX Matches 181; Conservative 1; Mismatches 10; Indels 5; Gaps 2;  
XX  
XX 1 MPGRKHQHFQEPBGVCGKYLFGFNIVFWLGLAFLAIGLWAGKGVLSNLSALTDLGG 60  
XX 1 MPGRKHQHFQEPBGVCGKYLFGFNIVFWLGLAFLAIGLWAGKGVLSNLSALTDLGG 60

Qy 61 LDPVWLFPVVVGWVSLGAGC---IGALRENTFLAKFFSVFLGLIFFLELATGILARVF 117  
Db 61 LDPVWL--VGSWRHRVGAGLCWAAGALRENTFLAKFFSVFLGLIFFLELATGILARVF 118  
Qy 118 KDWIRDLQNFNNNVKAYRDDIDQLNLIDFAQEWYSCCGARGPNDNLIYFNCTDLNP 177  
Db 119 KDWIRDLQNLFINNVKAYRDDIDQLNLIDFAQEWYSCCGARGPNDNLIYFNCTDLNP 178  
Qy 178 SRERCGVPFSCCVDPDA 194  
Db 179 SRERCGVPFSCCVDPDA 195  
RESULT 7  
AAB53438  
ID AAB53438 standard; Protein; 271 AA.  
XX AC AAB53438;  
XX 09-MAR-2001 (first entry)  
XX Human colon cancer antigen protein sequence SEQ ID NO:978.  
XX  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX identification; cytostatic; cardioactive; neuroprotective; vulnary;  
XX immunomodulatory; muscular; gynaecological; gastrointestinal;  
XX nephrotropic; antineoplastic; antibacterial; gene therapy; wound;  
XX neural disorder; immune system disorder; muscular disorder;  
XX reproductive disorder; gastrointestinal disorder; renal disorder;  
XX infectious disease; cardiovascular disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200055351-A1.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US05883.  
XX  
XX 12-MAR-1999; 99US-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM;  
XX  
XX WPI; 2000-587534/55.  
XX  
XX N-PSDB; AAC98195.  
XX  
XX Colon cancer associated gene sequences, referred to as colon cancer  
XX antigens, useful for the treatment, prevention, and diagnosis of colon  
XX disorders such as colon cancer -  
XX  
XX Claim 11; Page 1550-1551; 2104pp; English.  
XX  
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The  
XX human colon cancer antigens can have cytostatic, cardioactive, muscular;  
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
XX vulnary, nephrotropic, antineoplastic, antibacterial activities, and  
XX can be used in gene therapy. The colon cancer antigen polynucleotides,  
XX proteins and antibodies to the proteins are useful for the prevention,  
XX treatment and diagnosis of colon disorders, such as colon cancer. The  
XX polynucleotides may be used in diagnostics and research, such as for  
XX chromosome identification, and as hybridization probes. The proteins  
XX may also be used to prevent diseases such as neural disorders, immune  
XX system disorders, muscular disorders, reproductive disorders, infectious  
XX gastrointestinal disorders, wounds, renal disorders, infectious  
XX diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
XX AAB54007 represent sequences used in the exemplification of the present  
XX invention.  
XX  
XX Sequence 271 AA;

Query Match 69.0%; Score 885; DB 21; Length 271;  
 Best Local Similarity 79.9%; Pred. No. 2e-85;  
 Matches 155; Conservative 16; Mismatches 21; Indels 2; Gaps 1;

QY 1 MPKGHOFQPEVCCCKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNISALTDLGG 60  
 DB 4 MSCK--HYKGPVSCCKYFIFGFNFVFLGTLGIGLWANGKGVLSNISITDLCG 61

QY 61 LDPVWLFLVVGVMVGLFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDW 120  
 DB 62 FDPVWLFLVVGVMVGLFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDW 121

QY 121 IRDLNFFINNKKAYRDDIDLQNLIDFAQYVSCCGARGPNDWNLNIYFNCTDLNPSRE 180  
 DB 122 IKDLQYFFINNIRAYRDDIDLQNLIDFTQYVWQCCGAFGADDWNLNIYFNCTDSNASRE 181

QY 181 RCGVPFSCCVDRPA 194  
 DB 182 RCGVPFSCCTKDP A 195

RESULT 8  
 AAB49505  
 ID AAB49505 standard; Protein; 268 AA.  
 AC AAB49505;  
 DT 09-MAR-2001 (first entry)  
 DE Clone HE8EJ16 #1.  
 KW Human; 4 transmembrane superfamily receptor protein;  
 KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
 KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
 KW bacterial infection; viral; fungal.  
 XX Homo sapiens.  
 XX WO200070076-A1.  
 XX 23-NOV-2000.  
 XX 18-MAY-2000; 2000WO-US13504.  
 XX 19-MAY-1999; 99US-0135122.  
 XX 03-JUN-1999; 99US-0137797.  
 XX 11-JUN-1999; 99US-0138573.  
 XX 18-AUG-1999; 99US-0149447.  
 XX 28-JAN-2000; 2000US-0178770.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;  
 PI Rosen CA;  
 XX WPI: 2001-007502/01.  
 XX N-PSDB; AAC90015.  
 XX Isolated nucleic acid molecule encoding human soluble 4 transmembrane  
 PT superfamily receptor protein, useful for diagnosing, treating and/or  
 PT preventing disorders e.g. Alzheimer's, cancer and arrhythmia -  
 XX Claim 11; Page 282; 297pp; English.  
 XX The present invention relates to isolated nucleic acids and proteins  
 CC encoding human soluble 4 transmembrane superfamily receptor protein (see  
 CC AAC90012-C90023 and AAB49513). The present sequence is one such  
 CC protein. The present protein can be used to screen for binding partners  
 CC and molecules which modify its activity. Antibodies specific for the  
 CC present protein can be used to treat and/or prevent diseases associated  
 CC with aberrant expression or activity of the present protein e.g.  
 CC endocrine disorders e.g. Addison's disease, (cardio)vascular diseases

CC e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural  
 disorders e.g. Alzheimer's and Parkinson's disease, reproductive  
 CC disorders, skin disorders e.g. psoriasis, renal system disorders e.g.  
 nephritis, (auto)immune system disorders e.g. graft vs. host disease,  
 CC hyperproliferative disorders e.g. neoplasms of the pancreas, ocular  
 disorders e.g. glaucoma and infections caused by bacteria, viruses and  
 CC fungi.  
 XX Sequence 268 AA;  
 SQ

Query Match 61.7%; Score 791; DB 22; Length 268;  
 Best Local Similarity 72.2%; Pred. No. 1.8e-75;  
 Matches 140; Conservative 16; Mismatches 36; Indels 2; Gaps 1;

QY 1 MPKGHOFQPEVCCCKYFLFGFNIVFWVLGALFLAIGLWANGKGVLSNISALTDLGG 60  
 DB 1 MSCK--HYKGPVSCCKYFIFGFNFVFLGTLGIGLWANGKGVLSNISITDLCG 58

QY 61 LDPVWLFLVVGVMVGLFAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDW 120  
 DB 59 FDPVWLFLVVGVMVGLFAGCIGAYGKTLSPSSFFLCSWELSFLELATGILAFVFKDW 118

QY 121 IRDLNFFINNKKAYRDDIDLQNLIDFAQYVSCCGARGPNDWNLNIYFNCTDLNPSRE 180  
 DB 119 IKDLQYFFINNIRAYRDDIDLQNLIDFTQYVWQCCGAFGADDWNLNIYFNCTDSNASRE 178

QY 181 RCGVPFSCCVDRPA 194  
 DB 179 RCGVPFSCCTKDP A 192

RESULT 9  
 AAY42381  
 ID AAY42381 standard; Protein; 193 AA.  
 AC AAY42381;  
 XX 09-DEC-1999 (first entry)  
 DT Amino acid sequence of dk329\_1.  
 DE secreted protein; cDNA library; clone; transmembrane protein;  
 KW signal sequence cloning; hybridization cloning; gene therapy;  
 KW receptor.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 FH Peptide 71..83  
 FT Peptide /label= Leader/Signal peptide  
 FT Protein 84..193  
 FT /label= dk329\_1 mature protein  
 XX WO9942470-A1.  
 XX 26-AUG-1999.  
 XX 18-FEB-1999; 99WO-US03458.  
 XX 18-FEB-1998; 98US-0075038.  
 XX 17-FEB-1999; 99US-0251600.  
 XX (GENY) GENETICS INST INC.  
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Merberg D;  
 PI Treacy M, Agostino MJ, Steininger RJ;  
 XX WPI: 1999-518580/43.  
 XX N-PSDB; AAZ20854.  
 XX New polynucleotides encoding human secreted proteins used for  
 PT therapeutic, diagnostic and research purposes.

PS	Claim 13; Page 101-102; 125pp; English.	PD	11-OCT-2001.
XX	This is the amino acid sequence of the dk329_1 protein, which is derived from the dk329_1 clone isolated from a human fetal kidney cDNA library.	XX	22-MAR-2001; 2001WO-US09369.
CC	The PNs and proteins of the invention are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, hematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, and and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumor invasion suppressor activity, and tumor inhibition activity. The PNs are also stated to be useful for gene therapy. Other activities include inhibiting the growth, infection or function of bacteria, fungi, viruses and other parasites; effecting bodily characteristics such as, e.g. weight, color, skin, etc., effecting biorhythms or circadian cycles; enhancing fertility; treatment of depression; treatment of pain; hormonal or endocrine activity.	XX	30-MAR-2000; 2000US-0539330.
CC		PR	04-DEC-2000; 2000US-0729674.
CC		XX	(GENY ) GENETICS INST INC.
CC		PI	Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C; Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG; Clark H, Feghtel K, Merberg D;
CC		XX	WPI: 2001-639363/73.
CC		DR	N-PSDB; AAS59291.
CC		PT	Secreted human proteins, useful as vaccine for treating various diseases such as autoimmune disorders (e.g. multiple sclerosis), and nervous system disorders (e.g. stroke) -
CC		XX	Disclosure; Page 577; 619pp; English.
CC		CC	The invention relates to novel human secreted proteins, the nucleic acids encoding them. The protein may exhibit cytokine, cell proliferation or cell differentiation activity or may induce production of other cytokines in certain cell populations and may exhibit immune stimulating or immune suppressing activity, which is useful for the treatment of various immune deficiencies and disorders e.g. severe combined immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation. The proteins are also useful in the treatment of diseases and disorders including tissue, skin and organ transplantation and in graft-versus-host diseases (GVHD), in the induction of tumour immunity, myeloid or lymphoid cell deficiencies, wound healing and tissue repair, in the treatment of burns, incisions and ulcers; as well as in treatment of periodontal diseases, osteoporosis or osteoarthritis, mediated by inflammatory processes, diseases of the peripheral nervous system, Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections, infarction of cardiac and central nervous system vessel e.g. stroke, sepsis, inflammatory bowel disease, ulcers, bone regeneration. The protein, having activin- or inhibin-related activities is useful as a contraceptive based on the ability of inhibiting to decrease fertility in female mammals and decrease spermatogenesis in male mammals. The proteins and nucleic acids are also useful as food supplements. The present sequence represents a secreted protein of the invention.
XX		SQ	Sequence 193 AA;
XX		Query Match	54.5%; Score 699; DB 22; Length 193;
XX		Best Local Similarity	62.9%; Pred. No. 7.1e-66;
XX		Matches 139; Conservative 16; Mismatches 30; Indels 36; Gaps 4;	
QY	1 MPGKHQHFQEPBVGCGKVFLLFGFNIVFWLGAFLAIGLWAGKGVLSNLSALTDLGG 60	QY	1 MPGKHQHFQEPBVGCGKVFLLFGFNIVFWLGAFLAIGLWAGKGVLSNLSALTDLGG 60
DB	1 MSGK--HYKGPVSCCIKFIYFGFNIFWFLGITFLGIGLWAGKGVLSNLSITDLGG 58	DB	1 MSGK--HYKGPVSCCIKFIYFGFNIFWFLGITFLGIGLWAGKGVLSNLSITDLGG 58
QY	61 LDPVWLVVVGVMVILGPGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDW 120	QY	61 LDPVWLVVVGVMVILGPGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDW 120
DB	59 FDPVWLVVVGVMVILGPGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDW 118	DB	59 FDPVWLVVVGVMVILGPGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDW 118
QY	121 IRDQLNFFNNVKAIRYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRE 180	QY	121 IRDQLNFFNNVKAIRYRDDIDLQNLIDFAQYWSCCGARGPNDNLIYFNCTDLNPSRE 180
DB	119 IKDQLYFFNNIRAYRDDIDLQNLIDFTQY-----IPMQV 155	DB	119 IKDQLYFFNNIRAYRDDIDLQNLIDFTQY-----IPMQV 155
QY	181 RCGVPFSCVDRPA-----MSSTPSVAMSGSNWSWSR 214	QY	181 RCGVPFSCVDRPA-----MSSTPSVAMSGSNWSWSR 214
DB	156 ESDVAF----HSPAALKIPQKMSSTLSVAMMPGKNQKLTSR 192	DB	156 ESDVAF----HSPAALKIPQKMSSTLSVAMMPGKNQKLTSR 192
XX		OS	Homo sapiens.
XX		PN	WO200175068-A2.
XX		XX	RESULT 11
XX		XX	ABB55782
XX		XX	ID ABB55782 standard; Protein; 193 AA.

XX ABB55782;  
 XX 14-FEB-2002 (first entry)  
 XX Human polypeptide SEQ ID NO 170.  
 XX Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
 KW autoimmune disorder; bacterial infection; fungal infection; cancer; tumour;  
 KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
 KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
 KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnery;  
 KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
 KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
 KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory.  
 XX Homo sapiens.  
 XX US2001039335-A1.  
 XX 08-NOV-2001.  
 XX 04-DEC-2000; 2000US-0729674.  
 XX 10-APR-1997; 97US-126425P.  
 XX 04-DEC-1997; 97US-067454P.  
 XX 20-DEC-1997; 97US-068379P.  
 XX 02-JAN-1998; 98US-070346P.  
 XX 07-JAN-1998; 98US-070643P.  
 XX 08-JAN-1998; 98US-070755P.  
 XX 13-JAN-1998; 98US-071304P.  
 XX 22-JAN-1998; 98US-072134P.  
 XX 30-JAN-1998; 98US-073095P.  
 XX 18-FEB-1998; 98US-075038P.  
 XX 30-MAR-2000; 2000US-0539330.  
 XX 23-NOV-1998; 98US-0197886.  
 XX (JACO/) JACOBS K.  
 XX (MCCO/) MCCOY J M.  
 XX (LAVA/) LAVALLIE E R.  
 XX (COLL/) COLLINS-RACIE L A.  
 XX (EVAN/) EVANS C.  
 XX (MERB/) MERBERG D.  
 XX (TREA/) TREACY M.  
 XX (AGOS/) AGOSTINO M J.  
 XX (STEI/) STEININGER R J.  
 XX (SPAU/) SPAULDING V.  
 XX (WONG/) WONG G G.  
 XX (CLAR/) CLARK H.  
 XX (FECH/) FECHTEL K.  
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
 PI Wong GG, Clark H, Fechtel K;  
 XX WPI; 2002-040725/05.  
 XX N-PSDB; ABA90960.  
 XX New secreted proteins and encoding polynucleotides, useful in gene  
 PT therapies, particularly for preventing or treating autoimmune  
 PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
 PT stroke or inflammations  
 XX Disclosure; Page 307-308; 349pp; English.  
 XX The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
 CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially  
 CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
 CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
 CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
 CC are deposited with the American Type Culture Collection (ATCC) with  
 CC accession number 98599. The polynucleotides and encoded polypeptides have

CC cytostatic, anti-inflammatory, immunomodulator, vulnery;  
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
 CC and anti-inflammatory activity and acting as cytokine modulators,  
 CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
 CC suppressors. The polypeptides and polynucleotides are useful in gene  
 CC therapies, particularly for preventing, treating or ameliorating any of  
 CC the following diseases: immune deficiency and disorders; e.g. bacterial  
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
 CC osteoarthritis; central and peripheral nervous system diseases and  
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
 CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
 CC or systemic inflammatory response syndrome; ischaemia-reperfusion  
 CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
 CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
 CC foliaceus.  
 XX SQ Sequence 193 AA;  
 Query Match 54.5%; Score 699; DB 23; Length 193;  
 Best Local Similarity 62.9%; Pred. No. 7.1e-66;  
 Matches 139; Conservative 16; Mismatches 30; Indels 36; Gaps 4;  
 QY 1 MPCKQHFOEPVGGCGKFFLPGFNIVFWLQALFLATGLWAGKGVLSNLSALTDLGG 60  
 DB 1 MSGK--HYKGEVSCCIKFIYFEGNVIWFLGITFLGLWAGKGVLSNLSALTDLGG 58  
 QY 61 LDPVWLVVVGWVMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGLAFVFKDW 120  
 DB 59 FDPVWLVVVGWVMSVLGFAGCIGALRENTFLKFFSVFLGLIFLELATGLAFVFKDW 118  
 QY 121 IRDQLNFFINNKKAYRDDIDLQNLIDFAQYVSCCGARGPNWNLNINNYENCTDLNPSRE 180  
 DB 119 IKDQYFFINNIRAYRDDIDLQNLIDFTQY-----IPMOV 155  
 QY 181 RCGVFPSCCVDPDPA-----MSSTPSVAMSGSNWSWSR 214  
 DB 156 ESDVAF---HSPAALKIPQKMSSTLSVAMFPGKQKLTSR 192  
 RESULT 12  
 ABB98695  
 ID ABB98695 standard; Protein; 512 AA.  
 XX AC ABB98695;  
 XX 13-JUN-2003 (first entry)  
 XX DE Rat gamma-hydroxybutyrate receptor.  
 XX rat; gamma-hydroxybutyrate; anti-epileptic; anxiolytic;  
 KW antineurodegeneration; antipsychotic; brain; dopamine; opioid; GABA;  
 KW gamma-aminobutyric acid; diagnosis; epilepsy; anxiety; sleep disorder;  
 KW behavioral disorder; neurodegeneration; Parkinson's disease; psychosis;  
 KW schizophrenia.  
 XX Rattus rattus.  
 XX WO200078948-A2.  
 XX 28-DEC-2000.  
 XX 19-JUN-2000; 2000WO-FR01687.  
 XX 18-JUN-1999; 99FR-0007784.  
 XX (UTPA-) UNIV PASTEUR LOUIS.  
 XX Andriananpandry C, Maitre M;  
 XX WPI; 2001-091570/10.



DR N-PSDB; ACC44092.  
 XX  
 PT New rat brain gamma-hydroxybutyrate receptor and its encoding nucleic  
 PT acid, useful for identifying agents for treating e.g. epilepsy -  
 XX  
 PS Claim 1; Page 62-63; 66pp; French.  
 XX  
 CC This sequence represents a novel rat gamma-hydroxybutyrate receptor  
 CC (GHR) isolated from rat brain hypocampal cell cDNA library. The  
 CC invention relates to the isolation of this novel sequence, fragments of  
 CC it and homologues of the sequence except for those homologues that having  
 CC GenBank accession numbers AAC 17120 (human tetraspan NET-4), AA615405,  
 CC AA967250 or A1467230. GHR is the receptor for gamma-hydroxybutyrate in  
 CC the rat brain and is involved in regulating dopaminergic, opioid and  
 CC GABA(gamma-aminobutyric acid)ergic activities. The nucleic acid that  
 CC encodes the protein is used: (1) as primers or probes for detection and  
 CC amplification, particularly for screening gene libraries to identify  
 CC promoters and regulators of the GHR gene; (2) for expression of  
 CC recombinant polypeptides; and (3) to detect allelic variants, mutations,  
 CC deletions, loss of heterozygosity or genetic abnormalities in the GHR  
 CC gene (for diagnosing diseases, or susceptibility, associated with  
 CC abnormal expression of GHR). The protein (or cells and transgenic  
 CC animals expressing it) is used: (1) to screen for agents that interact  
 CC with GHR; (2) to study expression/activity of the receptor, including  
 CC its interaction with other compounds; and (3) to raise antibodies (Ab)  
 CC specific for GHR. The Ab are used to detect/measure GHR in diagnostic  
 CC immunoassays. The agents are used to prevent or treat diseases associated  
 CC with abnormal expression or activity of GHR, particularly those  
 CC involving cerebral GABA(gamma-aminobutyric acid)ergic and/or dopaminergic  
 CC activities, e.g. epilepsy, anxiety, sleep or behavioral disorders,  
 CC withdrawal from addictive drugs, neurodegeneration (Parkinson's disease),  
 CC psychoses, schizophrenia and regulation of secretion of hormones (growth  
 CC hormone and prolactin) that are under dopaminergic control.  
 XX  
 SQ Sequence 512 AA;

Query Match 49.9%; Score 640.5; DB 22; Length 512;  
 Best Local Similarity 89.6%; Pred. No. 3.9e-59;  
 Matches 120; Conservative 3; Mismatches 10; Indels 1; Gaps 1;  
 QY 59 GGLDPVWLVVVGWMSVLGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVK 118  
 DB 97 GGLDPVWLVVVGWMSVLGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVK 156  
 QY 119 DWLRDQLNLFNNKAYRDDIDLQNLIDFAQEWSCCGARGPNDWNLIYFN-CTDLNP 177  
 DB 157 DWLRDQLNLFNNKAYRDDIDLQNLIDFAQEWSCCGARGPNDWNLIYFN-CTDLNP 216  
 QY 178 SRRCGVPPFSCVVR 191  
 DB 217 SRRCGVPPFSCVVR 230

RESULT 13  
 AA76266  
 ID AAY76266 standard; Protein; 270 AA.

AC AAY76266.  
 XX  
 DT 23-MAR-2000 (first entry)  
 XX  
 DE Human secreted protein encoded by gene 10 fragment.  
 DE  
 KW Human; secreted protein; cancer; tumour; developmental abnormality;  
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
 KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;  
 KW therapy; chromosome 10.  
 XX  
 OS Homo sapiens.  
 XX

PN W09958660-A1.  
 XX  
 PD 18-NOV-1999.  
 XX  
 PF 06-MAY-1999; 99WO-US09847.  
 XX  
 PR 12-MAY-1998; 98US-0085093.  
 PR 12-MAY-1998; 98US-0085094.  
 PR 12-MAY-1998; 98US-0085105.  
 PR 12-MAY-1998; 98US-0085180.  
 PR 18-MAY-1998; 98US-0085906.  
 PR 18-MAY-1998; 98US-0085920.  
 PR 18-MAY-1998; 98US-0085921.  
 PR 18-MAY-1998; 98US-0085922.  
 PR 18-MAY-1998; 98US-0085923.  
 PR 18-MAY-1998; 98US-0085924.  
 PR 18-MAY-1998; 98US-0085928.  
 PR 18-MAY-1998; 98US-0085925.  
 PR 18-MAY-1998; 98US-0085927.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Florence K, Ni J, Rosen CA, Carter KC, Moore PA;  
 PI Olsen HS, Shi Y, Young PE, Wei F, Brewer LA, Soppet DR;  
 PI Lafleur DW, Endress GA, Ebner R;  
 XX  
 DR WPI; 2000-062296/05.  
 DR N-PSDB; AAZ65350.  
 XX  
 CC New isolated human genes and the secreted polypeptides they encode,  
 CC useful for diagnosis and treatment of e.g. cancer, neurological  
 CC disorders, immune diseases, inflammation or blood disorders -  
 XX  
 PS Disclosure; Page 428-429; 475pp; English.

XX  
 CC AAZ65250 to AAZ65350 represent 97 isolated human secreted protein genes.  
 CC AAY76124 to AAY76223 are the secreted proteins encoded by the 97 human  
 CC genes. The gene encoding this protein was found to be on chromosome 10.  
 CC The genes and their corresponding secreted polypeptides are  
 CC useful for preventing, treating or ameliorating medical conditions,  
 CC e.g. by protein or gene therapy. Also pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new genes. Specific  
 CC uses are described for each of the 97 genes, based on which tissues they  
 CC are most highly expressed in, and include developing products for the  
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
 CC and foetal deficiencies, blood disorders, diseases of the immune system,  
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
 CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
 CC disorders, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The sequences shown in AAY76224 to AAY76424 represent fragments of the  
 CC secreted proteins.

XX  
 SQ Sequence 270 AA;

Query Match 49.8%; Score 638.5; DB 21; Length 270;  
 Best Local Similarity 49.4%; Pred. No. 2.9e-59;  
 Matches 119; Conservative 37; Mismatches 66; Indels 19; Gaps 2;  
 QY 5 HQHFQPEVGVCCCKYFLFGFNIVFWVLGALFLAIGLWAGCEKGVLSNISALTDLGGIDPV 64  
 DB 3 YYRSNAKVSCKWYKLLFSYNIIFWLAGVFLGVGLWAWSEKGVLSLTIVTRMHGIDPV 62  
 QY 65 WLPVVVGVWMSVLGAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDWRDQ 124  
 DB 63 VLVLWGVVWNTLGFAGCIGALRENTFLKFFSVFLGLIFLELATGILAFVKDWRDQ 122  
 QY 125 LNFIFNNKAYRDDIDLQNLIDFAQEWSCCGARGPNDWNLIYFNCTDLNPSRRCGV 184  
 DB 123 PREFFESNIKSYRDDIDLQNLIDSLQANQCCAGYGDWDNLNVCSCASYSREKCGV 182



QY 185 PFSCVDRPAMSTPS-----VAMMSGSNW-----SWSRAPYTPKAYWAS 225  
 DB 183 PFSCVDRPAMSTPS-----VAMMSGSNW-----SWSRAPYTPKAYWAS 242  
 QY 226 L 226  
 DB 243 I 243

RESULT 14  
 ID AAB87034 standard; Protein; 270 AA.  
 XX  
 AC AAB87034;  
 XX  
 DT 04-MAY-2001 (first entry)  
 XX  
 DE Human secreted protein TANGO 339, SEQ ID NO:3.  
 XX  
 KW Secreted protein; transmembrane protein; TANGO; human; drug screening;  
 KW activity modulator; expression modulator; cancer; immunological disorder;  
 KW cytosolic; immunomodulatory; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200109162-A2.  
 XX  
 PD 08-FEB-2001.  
 XX  
 PF 31-JUL-2000; 2000WO-US20935.  
 XX  
 PR 30-JUL-1999; 99US-0365164.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Fraser CC, Sharp JD, Kirst SJ, Barnes TM, Wrighton N, Myers PS;  
 PI Pan Y;  
 XX  
 DR WPI; 2001-138647/14.  
 XX  
 DR N-PSDB; AAF90628, AAF90629.  
 XX  
 PT Nucleic acids encoding secreted TANGO and MANGO polypeptides, useful  
 PT for the prevention, diagnosis and treatment of, e.g. cancers and immune  
 PT disorders -  
 XX  
 PS Claim 9; Page 220-221; 332pp; English.  
 XX  
 CC The invention relates to novel secreted/transmembrane proteins, and  
 CC nucleic acids encoding them. The novel proteins are designated TANGO 339,  
 CC TANGO 353, TANGO 358, TANGO 365, TANGO 368, TANGO 369, TANGO 383, TANGO  
 CC 393, TANGO 402, MANGO 346 and MANGO 349 and are of human origin, and a  
 CC murine TANGO 393 is also included within the scope of the invention. The  
 CC invention also encompasses fragments and variants of the proteins of the  
 CC invention, and nucleic acids encoding them. The invention additionally  
 CC relates to host cells comprising a nucleic acid of the invention; methods  
 CC for the production of a protein of the invention; an antibody specific  
 CC for a protein of the invention; methods for detecting a protein or  
 CC nucleic acid of the invention; and methods of identifying agents which  
 CC bind to or modulate the activity of a protein of the invention. The novel  
 CC secreted proteins, nucleic acids encoding them, and antibodies against  
 CC them may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate expression or activity of the secreted  
 CC proteins. The secreted proteins of the invention may also be used to  
 CC identify modulators of expression or activity, which may be useful in  
 CC the treatment of disorders associated with the proteins of the  
 CC invention e.g., cancers and immunological disorders. The present  
 CC sequence represents a human TANGO protein of the invention.

QY Sequence 270 AA;  
 SQ

Query Match 49.8%; Score 638.5; DB 22; Length 270;  
 Best Local Similarity 49.4%; Pred. No. 2.9e-59;  
 Matches 119; Conservative 37; Mismatches 66; Indels 19; Gaps 2;

QY 5 HQHFOEPEVCCGKYFLFGFNIVFVNLGALFLATGLWAGKEGVLNLSALTDLGLDPV 64  
 DB 3 YYRYSNAKVCWYKYLFLFSYNIIFWLAGVFLGVLWAWSEKGVLSDLTKVTRMHGIDPV 62  
 QY 65 WLFVWVGVMVSLGFCAGCIGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDWIRDO 124  
 DB 63 VLVLVGVVMTLGFAGCVGALRENTFLKFFSVFLGLIFFLELATGILAFVFKDWIRDO 122  
 QY 125 LNFFINNKKAYRDDIDLQNLIDPAQYWSCCGARGNDNMLNINYNCTDLNPSRCGV 184  
 DB 123 PREFFESNIKSYRDDIDLQNLIDSLQKQANCCGAYGPEMDLNVYFNCSGASYSREKGV 182  
 QY 185 PFSCVDRPAMSTPS-----VAMMSGSNW-----SWSRAPYTPKAYWAS 225  
 DB 183 PFSCVDRPAMSTPS-----VAMMSGSNW-----SWSRAPYTPKAYWAS 242  
 QY 226 L 226  
 DB 243 I 243

RESULT 15  
 ID AAE18535 standard; Protein; 270 AA.  
 XX  
 AC AAE18535;  
 XX  
 DT 17-MAY-2002 (first entry)  
 XX  
 DE Human tetraspan protein, TSPAN-7.  
 XX  
 KW Human; tetraspan protein; TSPAN-7; hyperproliferative disorder;  
 KW neoplastic disease; prostate cancer; antisense therapy; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200206340-A2.  
 XX  
 PD 24-JAN-2002.  
 XX  
 PF 13-JUL-2001; 2001WO-US22134.  
 XX  
 PR 14-JUL-2000; 2000US-218280P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Reinhard CJ, Garcia PD;  
 XX  
 DR WPI; 2002-179783/23.  
 XX  
 DR N-PSDB; AAD29623.  
 XX  
 PT New inhibitor of tetraspan protein useful for reducing the expression  
 PT or activities of tetraspan protein in cells, and for treating a  
 PT hyperproliferative disorder, especially cancer in a mammal -  
 XX  
 PS Claim 17; Fig 3; 80pp; English.  
 XX  
 CC The invention relates to inhibitors of tetraspan protein, TSPAN-7  
 CC which are used to reduce or decrease the expression of TSPAN-7  
 CC in a mammalian cell and for treating hyperproliferative disorder,  
 CC especially cancer in a mammal. The invention also provides TSPAN-7  
 CC polypeptides and polynucleotides. TSPAN-7 inhibitors are useful for  
 CC inhibiting proliferation of mammalian cells, including tumour cells,  
 CC for decreasing the side effects of cancer therapy and for treating  
 CC neoplastic diseases. They are used to modulate TSPAN-7 expression  
 CC and function in cancer cells, particularly in prostate cancer cells.  
 CC They are also used in antisense therapy. The present sequence is  
 CC human TSPAN-7 protein.

QY Sequence 270 AA;  
 SQ

Query Match 49.8%; Score 638.5; DB 23; Length 270;

Search completed: November 21, 2003, 13:43:36  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2003, 15:29:38 ; Search time 335 Seconds  
(without alignments)  
2273.109 Million cell updates/sec

Title: US-09-972-970-4

Perfect score: 1283

Sequence: 1 MFGKHQHFQEPVGGCGKYF.....RAPYTPKAYWASLRSGRTT 233

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cn2\_1/USPTO.spool/US09972970/runat\_21112003.125336.27261/app\_query.fasta\_1.391  
-DB=Published Applications\_NA -QWTF=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USRP=US09972970 @CGN 1.1 221 @runat\_21112003.125336.27261  
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:

1: /cn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cn2\_6/ptodata/1/pubpna/US07\_PUB.seq.\*  
3: /cn2\_6/ptodata/1/pubpna/US06\_PUB.seq.\*  
4: /cn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cn2\_6/ptodata/1/pubpna/US07\_NEW\_PUBCOMB.seq.\*  
7: /cn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
8: /cn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
9: /cn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
10: /cn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
12: /cn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cn2\_6/ptodata/1/pubpna/US10\_NEW\_PUBCOMB.seq.\*  
17: /cn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1283	100.0	2538	10	US-09-972-970-2 Sequence 2, Appli

2	1215.5	94.7	813	10	US-09-934-268-3	Sequence 3, Appli
3	1215.5	94.7	813	14	US-10-162-435-37	Sequence 37, Appl
4	1215.5	94.7	2461	12	US-10-252-157-153	Sequence 153, App
5	1215.5	94.7	3184	10	US-09-934-268-1	Sequence 1, Appli
6	1215.5	94.7	3184	14	US-10-162-435-35	Sequence 35, Appl
7	1212.5	94.5	1126	14	US-10-103-196-4	Sequence 4, Appli
8	958.5	74.7	1655	9	US-09-925-299-205	Sequence 205, App
9	958.5	74.7	1655	11	US-09-925-299-205	Sequence 205, App
10	918.5	71.6	1174	14	US-10-103-196-5	Sequence 5, Appli
11	892.5	69.6	1178	14	US-10-103-196-10	Sequence 10, Appl
12	807	62.9	1110	9	US-09-729-674-169	Sequence 169, App
13	801.5	62.5	1988	11	US-09-822-846-126	Sequence 126, App
14	790	61.6	864	10	US-09-875-440-1	Sequence 1, Appli
15	654.5	51.0	1368	11	US-09-905-670-1	Sequence 1, Appli
16	654.5	51.0	2715	11	US-09-796-753-125	Sequence 125, App
17	641.5	50.0	2672	11	US-09-892-877-20	Sequence 20, Appl
18	641.5	50.0	2672	11	US-09-948-783-20	Sequence 20, Appl
19	635.5	49.5	1428	11	US-09-991-053-15	Sequence 15, Appl
20	635.5	49.5	1428	11	US-09-957-187-15	Sequence 15, Appl
21	583.5	45.5	1667	11	US-09-892-877-266	Sequence 266, App
22	583.5	45.5	1667	11	US-09-948-783-288	Sequence 288, App
23	537	41.9	368	13	US-10-042-417-37	Sequence 37, Appl
24	490.5	38.2	839	14	US-10-106-698-2019	Sequence 2019, Ap
25	476.5	37.1	1932	11	US-09-984-271-41	Sequence 41, Appl
26	427.5	33.3	816	11	US-09-991-053-78	Sequence 78, Appl
27	427.5	33.3	816	11	US-09-957-187-78	Sequence 78, Appl
28	334.5	26.1	1778	11	US-09-946-374-122	Sequence 122, App
29	334.5	26.1	1778	12	US-10-015-387A-122	Sequence 122, App
30	334.5	26.1	1778	12	US-10-006-130A-122	Sequence 122, App
31	334.5	26.1	1778	12	US-10-199-672-277	Sequence 277, App
32	334.5	26.1	1778	12	US-10-006-172A-122	Sequence 122, App
33	334.5	26.1	1778	12	US-10-187-749-277	Sequence 277, App
34	334.5	26.1	1778	12	US-10-194-457-277	Sequence 277, App
35	334.5	26.1	1778	12	US-10-184-642-277	Sequence 277, App
36	334.5	26.1	1778	12	US-10-196-747-277	Sequence 277, App
37	334.5	26.1	1778	12	US-10-015-392A-122	Sequence 122, App
38	334.5	26.1	1778	12	US-10-017-253A-122	Sequence 122, App
39	334.5	26.1	1778	12	US-10-173-689-277	Sequence 277, App
40	334.5	26.1	1778	12	US-10-173-690-277	Sequence 277, App
41	334.5	26.1	1778	12	US-10-173-691-277	Sequence 277, App
42	334.5	26.1	1778	12	US-10-173-692-277	Sequence 277, App
43	334.5	26.1	1778	12	US-10-173-694-277	Sequence 277, App
44	334.5	26.1	1778	12	US-10-173-698-277	Sequence 277, App
45	334.5	26.1	1778	12	US-10-173-699-277	Sequence 277, App

## ALIGNMENTS

RESULT 1

US-09-972-970-2

Sequence 2, Application US/09972970

Patent No. US2002016493A1

GENERAL INFORMATION:

APPLICANT: Shi et al.

TITLE OF INVENTION: TMSF Receptor Polynucleotides, Polypeptides, and Antibodies

FILE REFERENCE: PT056pi

CURRENT APPLICATION NUMBER: US/09/972,970

PRIOR FILING DATE: 2001-10-10

PRIOR APPLICATION NUMBER: PCT/US01/11130

PRIOR FILING DATE: 2001-04-05

PRIOR APPLICATION NUMBER: 60/195,336

PRIOR FILING DATE: 2000-04-10

NUMBER OF SEQ ID NOS: 8

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 2

LENGTH: 2538

TYPE: DNA

ORGANISM: Homo sapiens

US-09-972-970-2

Alignment Scores: 2.59e-139 Length: 2538  
Pred. No.: 1283.00 Matches: 233  
Score:

Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-972-970-4 (1-233) x US-09-972-970-2 (1-2538)

QY 1 MetProGlyValHisGlnHisPheGlnGluProGluValGlyCysCysGlyValPhe 20  
DB 119 ATCCGGGAGGACACAGGACATTCAGGAACCCGAGGTGGCTCGGGAATATCTTC 178

QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
DB 179 CTGTTTGGCTTCAACATTTCTGGTGTCTGGAGCCCTGTTCTTGGCCATCGGCTC 238

QY 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
DB 239 TGGGCCCTGGGGTGAGAGGGTGTCTCTCAACATCTCTCGCTGACCGATCTGGGAGGC 298

QY 61 LeuAspProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80  
DB 299 CTGACCTGTGTGGCTGTTTGTAGTGTGGAGGCGTCATGTCGGCTGGGCTTGGCC 358

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
DB 359 GGCTGCATCGGGCTCTCCGGAGAACACATTTCTCTGCTCAAGTTTTCTCAGTGTCTT 418

QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
DB 419 GGCTCATCTCTTCTCGGAGCTGGCAACAGGATCTTGGCTTCGTATTTCAAGGACTGG 478

QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaValArgAspAspIle 140  
DB 479 ATTCGAGACAGCTCAATTTCTTCAITTAACAACAGTCAAGCCATCCGGATGACAT 538

QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysCysGlyAlaArgGly 160  
DB 539 GACCTCAGAACCTCATTTGCTTCAGGAATATTGGTCTTGGCTGGGAGCCCGAGGG 598

QY 161 ProAsnAspTrpAsnLeuAsnIleTyrrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
DB 599 CCTAATGACTGGAACTCAATATCTATTCAACTGCATGACTTGAACCCGAGCCGAGAG 658

QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrProSer 200  
DB 659 CGCTGGGGTGCCCTTCTCTGCTGTGTGAGGACCCCTCGATGTCTTCAACACCCAGT 718

QY 201 ValAlaMetMetSerGlySerAsnTrpSerTrpSerSetArgAlaProTyrrProLys 220  
DB 719 GTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGCTCCATACACCAAG 778

QY 221 AlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
DB 779 GCTGTGTGGCCAGTTTGAGAGTGGCTGCAGACAACC 817

## RESULT 2

US-09-934-268-3  
; Sequence 3, Application US/09934268  
; Patent No. US20020172986A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY  
; FILE REFERENCE: 10448-079001  
; CURRENT APPLICATION NUMBER: US/09/934,268  
; CURRENT FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: 60/226,612  
; PRIOR FILING DATE: 2000-08-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 813  
; TYPE: DNA

ORGANISM: Homo sapiens  
US-09-934-268-3

Alignment Scores:  
Pred. No.: 3,73e-132 Length: 813  
Score: 1215.50 Matches: 227  
Percent Similarity: 96.60% Conservatives: 0  
Best Local Similarity: 96.60% Mismatches: 6  
Query Match: 94.74% Indels: 2  
DB: 10 Gaps: 1

US-09-972-970-4 (1-233) x US-09-934-268-3 (1-813)

QY 1 MetProGlyValHisGlnHisPheGlnGluProGluValGlyCysCysGlyValPhe 20  
DB 1 ATCCCGGAGGACACAGGACATTTCCAGGAACCTTGAGGTGGCTCGGGAATATCTTC 60

QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
DB 61 CTGTTTGGCTTCAACATTTGTCTTCTGGTGTCTGGAGCCCTGTTCTTGGCTATCGGCTC 120

QY 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
DB 121 TGGGCCCTGGGGTGAGAGGGTGTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 180

QY 61 LeuAspProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80  
DB 181 CTTGACCCGCTGTGGCTGTTTGTGTGTAGTTGGAGGCGTCATGTCTGGTGTCTGGCTT 240

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
DB 241 GGCTGCATTTGGGCCCCCTCGGAGAACACCTTCTCTCAAGTTTTTCTCCGTGTTCTC 300

QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
DB 301 GGCTCATCTTCTTCTCGAGCTGGCAACAGGATCTTGGCCCTTGTCTTCAAGGACTGG 360

QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaValArgAspAspIle 140  
DB 361 ATTCGAGACAGCTCAACCTTCTTCAATCAACAACAGCTCAAGCCCTACCGGAGCAGCAT 420

QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysCysGlyAlaArgGly 160  
DB 421 GACCTCAGAACCTCATTTGCTTCAGGAATATCTGCTTGTGTGGAGCCCGAGGC 480

QY 161 ProAsnAspTrpAsnLeuAsnIleTyrrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
DB 481 CCCAATGACTGGAACTCAATATCTACTTCAACTGCATGACTTGAACCCCGAGGAG 540

QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrPr 199  
DB 541 CGCTGGGGTGCCCTTCTCTGCTGCTCGCTCAGGAGCCCTCGGAGGATGTCTCAACACC 600

QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrrPr 219  
DB 601 CAGTGTGGCTAGCAGCTCGGCTCAAACTGGAGCTGGAGCAGCAGGCTTCTCCACACC 660

QY 219 olysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
DB 661 AAAGGCTGGTGGCCAGTTTTCAGAGTGGCTGCAGGACAACC 703

## RESULT 3

US-10-162-435-37  
; Sequence 37, Application US/10162435  
; Publication No. US20030096305A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Glucksman, Maria Alexandra  
; APPLICANT: Curtis, Rory A. J.  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Bandaru, Rajasekhar  
; APPLICANT: Leiby, Kevin R.  
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND

;; TITLE OF INVENTION: CELL SURFACE PROTEIN FAMILY MEMBERS

;; FILE REFERENCE: 10448-189001

;; CURRENT APPLICATION NUMBER: US/10/162,435

;; CURRENT FILING DATE: 2002-06-04

;; PRIOR APPLICATION NUMBER: US 09/836,499

;; PRIOR FILING DATE: 2001-04-17

;; PRIOR APPLICATION NUMBER: PCT/US01/12420

;; PRIOR FILING DATE: 2001-04-17

;; PRIOR APPLICATION NUMBER: US 60/197,507

;; PRIOR FILING DATE: 2000-04-18

;; PRIOR APPLICATION NUMBER: US 09/891,008

;; PRIOR FILING DATE: 2001-06-25

;; PRIOR APPLICATION NUMBER: PCT/US01/19963

;; PRIOR FILING DATE: 2001-06-25

;; PRIOR APPLICATION NUMBER: US 60/214,220

;; PRIOR FILING DATE: 2000-06-23

;; PRIOR APPLICATION NUMBER: US 09/860,868

;; PRIOR FILING DATE: 2001-05-18

;; PRIOR APPLICATION NUMBER: PCT/US01/16013

;; PRIOR FILING DATE: 2001-05-18

;; PRIOR APPLICATION NUMBER: US 60/205,674

;; PRIOR FILING DATE: 2000-05-19

;; PRIOR APPLICATION NUMBER: US 09/886,429

;; PRIOR FILING DATE: 2001-06-21

;; PRIOR APPLICATION NUMBER: PCT/US01/20055

;; PRIOR FILING DATE: 2001-06-21

;; PRIOR APPLICATION NUMBER: US 60/213,963

;; PRIOR FILING DATE: 2000-06-23

;; PRIOR APPLICATION NUMBER: US 10/041,406

;; PRIOR FILING DATE: 2002-01-08

;; PRIOR APPLICATION NUMBER: PCT/US02/00275

;; PRIOR FILING DATE: 2002-01-08

;; PRIOR APPLICATION NUMBER: US 60/260,286

;; PRIOR FILING DATE: 2001-01-08

;; PRIOR APPLICATION NUMBER: US 09/934,268

;; PRIOR FILING DATE: 2001-08-21

;; PRIOR APPLICATION NUMBER: PCT/US01/41811

;; PRIOR FILING DATE: 2001-08-21

;; PRIOR APPLICATION NUMBER: US 60/226,612

;; PRIOR FILING DATE: 2000-08-21

;; NUMBER OF SEQ ID NOS: 38

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 37

;; LENGTH: 813

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; US-10-162-435-37

Alignment Scores:

Pred. No.:

Score: 3.73e-132 Length: 813

Percent Similarity: 1215.50 Matches: 227

Best Local Similarity: 96.60% Conservative: 0

Query Match: 94.74% Mismatches: 6

DB: 14 Indels: 2

Gaps: 1

US-09-972-970-4 (1-233) x US-10-162-435-37 (1-813)

Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20

Db 1 ATGCCCGGCAAGCAGCATTTCCAGAACTTCAGAGTCTCGCGGAAATACTTC 60

Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40

Db 61 CTGTTTGGCTTCAACATTTCTCTGGGTGTGGAGCCCTGTTCTCGGCTATCGGCCTC 120

Qy 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60

Db 121 TGGGCTCTGGGTGAGAGGGGCTTCTCTCGAACATCTCAGGCTGACAGATCTCGGAGGC 180

Qy 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80

Db 181 CTTGACCCCGGTGGCTGTTTGTGTAGTTGGAGGCGTCATGTCGCTGGGCTTTGCT 240

Qy 81 GlyCysIleGlyValAlaLeuArgGluAenThrPheLeuLeuLysPhePheSerValPheLeu 100

Db 241 GGCTGCAITGGGGCCCTCCGGAGAACACCTTCTCTGCTCAAGTTTTTCTCCGTTTCCTC 300

Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTyr 120

Db 301 GGCTCACTTCTTCTCTGGAGCTGGCAACAGGGATCTCGCCCTTTGTCTTCAAGGACTGG 360

Qy 121 IleArgAspGlnLeuAsnPheIleAsnAenValLysAlaTyrArgAspAspIle 140

Db 361 ATTCGAGACCCAGCTCAACCTCTTCAACAACAACAGCTCAAGGCTTACCGGGACGACATT 420

Qy 141 AspLeuGlnAenLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160

Db 421 GACCTCCAGAACCTCATTGACTTCTCAGGAATATCTGCTGTGGAGCCCGAGGC 480

Qy 161 ProAsnAspTyrAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180

Db 481 CCCAATGACTGGAACTCAATATCTTCAACTGCACTGACTTGAACCCCGAGCGGAG 540

Qy 181 ArgCysGlyValPropheSerCysCysValArgAspProAla----MetSerSerThrPr 199

Db 541 CGCTCGGGGTGCCCCCTTCTCTGCTGCGTCAGGGACCCCTGCGGAGGATGCTCCTCAACACC 600

Qy 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219

Db 601 CAGTGTGGCTAGCAGCTCCGGCTCAAACTGGAGCTGGAGCAGCAGCGGCTTTCACACACC 660

Qy 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233

Db 661 AAAGCTCGGTGGGCCAGTTTGAGAACTGGCTGCGAGGACACC 703

RESULT 4

US-10-252-157-153

;; Sequence 153, Application US/10252157

;; Publication No. US20030190640A1

;; GENERAL INFORMATION:

;; APPLICANT: Faris, Mary

;; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER

;; FILE REFERENCE: PA-0027-1 US

;; CURRENT APPLICATION NUMBER: US/10/252,157

;; CURRENT FILING DATE: 2002-10-01

;; PRIOR APPLICATION NUMBER: 60/295,048

;; PRIOR FILING DATE: 2001-05-31

;; NUMBER OF SEQ ID NOS: 501

;; SOFTWARE: PERL Program

;; SEQ ID NO 153

;; LENGTH: 2461

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; NAME/KEY: misc feature

;; OTHER INFORMATION: Incyte ID No. US20030190640A1 234056.5

US-10-252-157-153

Alignment Scores:

Pred. No.: 1.89e-131 Length: 2461

Score: 1215.50 Matches: 227

Percent Similarity: 96.60% Conservative: 0

Best Local Similarity: 96.60% Mismatches: 6

Query Match: 94.74% Indels: 2

DB: 12 Gaps: 1

US-09-972-970-4 (1-233) x US-10-252-157-153 (1-2461)

Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20

Db 114 ATGCCCGGCAAGCAGCATTTCCAGAACTTCAGAGTCTCGCGGAAATACTTC 173

Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40

174 CTGTTTGGCTTCAACATTGCTTCTGGTCTGGAGCCCTGTTCTCGCTATCGGCCTC 233  
QY 41 TtpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 234 TGGCCCTGGGTGAGAGGGCGTCTCTCGAATCTCTAGCGCTGACAGATCTGGGAGGC 293  
QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 294 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGCGCTCATGTCGGTGTGGGCTTTGCT 353  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 354 GGTGTCATTGGGGCCCTCGGGAGAACCTTCTCTCAAGTCTTCTCGGTGTTCCCTC 413  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 414 GGTCTCATCTTCTCTGGAGTGGCAACAGGATCTCTGGCTTGTCTTCAAGGACTGG 473  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaLysArgAspIle 140  
Db 474 ATTCGAGACAGCTCAACCTTCTCATCAACAACAGCTCAAGGCTTACCGGACGACATT 533  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160  
Db 534 GACCTCCAGAACCTCATTTGCTCAGGAATACTGGTCTTGTCTGGAGCCCGAGGC 593  
QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 594 CCCATGACTGNAACCTCAATATCTACTTCACTGACACTGACTTGAACCCAGCCGGAG 653  
QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199  
Db 654 CGCTCGGGGTGCCCTTCTCTGCTGCGTCAGGACCCCTGCGGAGGATGCTCTCAACACC 713  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219  
Db 714 CAGTGTGGCTACGACGTCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCTCACACACC 773  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
Db 774 AAAGGCTGCGTGGCCAGCTTTCAGAAAGTGGCTGCAGGACCAACC 816

RESULT 5  
US-09-934-268-1  
; Sequence 1, Application US/09934268  
; Patent No. US20020172986A1  
; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; TITLE OF INVENTION: 23228, A NOVEL HUMAN TETRASPANIN FAMILY  
; FILE REFERENCE: 10448-079001  
; CURRENT APPLICATION NUMBER: US/09/934,268  
; CURRENT FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: 60/226,612  
; PRIOR FILING DATE: 2000-08-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3184  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (168)...(977)  
US-09-934-268-1

Alignment Scores:  
Pred. No.: 2,75E-131 Length: 3184  
Score: 1215.50 Matches: 227  
Percent Similarity: 96.60% Conservative: 0  
Best Local Similarity: 96.60% Mismatches: 6  
Query Match: 94.74% Indels: 2  
DB: 10 Gaps: 1

US-09-972-970-4 (1-233) x US-09-934-268-1 (1-3184)  
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20  
Db 168 ATGCCCGGCAAGCACCGCATTTCCAGGAACCTTGAGTCTGGCTGCGGAAATACTTTC 227  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 228 CTGTTTGGCTTCAACATTGCTTCTCTGGGTGCTGGAGCCCTGTTCTCGCTATTCGGCCCTC 287  
QY 41 TtpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 288 TGGGCTTGGGTGAGAGGGCGTCTCTCGAATCTCTAGCGCTGACAGATCTGGGAGGC 347  
QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 348 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGCGCTCATGTCGGTGTGGGCTTTGCT 407  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 408 GGTGTCATTGGGGCCCTCGGGAGAACCTTCTCTCAAGTCTTCTCGGTGTTCCCTC 467  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 468 GGTCTCATCTTCTCTGGAGTGGCAACAGGATCTCTGGCTTGTCTTCAAGGACTGG 527  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaLysArgAspIle 140  
Db 528 ATTCGAGACAGCTCAACCTTCTCATCAACAACAGCTCAAGGCTTACCGGACGACATT 587  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160  
Db 588 GACCTCCAGAACCTCATTTGCTCAGGAATACTGGTCTTGTCTGGAGCCCGAGGC 647  
QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 648 CCCATGACTGGAACCTCAATATCTACTTCACTGACACTGACTTGAACCCAGCCGGAG 707  
QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199  
Db 708 CGCTCGGGGTGCCCTTCTCTGCTGCGTCAGGACCCCTGCGGAGGATGTCCTCAACACC 767  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219  
Db 768 CAGTGTGGCTACGACGTCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTCTCACACACC 827  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
Db 828 AAAGGCTGCGTGGCCAGCTTTCAGAAAGTGGCTGCAGGACCAACC 870

RESULT 6  
US-10-162-435-35  
; Sequence 35, Application US/10162435  
; Publication No. US20030096305A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Glucksmann, Maria Alexandra  
; APPLICANT: Curtis, Rory A. J.  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Bandaru, Rajasekhar  
; APPLICANT: Leiby, Kevin R.  
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE-ASSOCIATED PROTEIN AND  
; FILE REFERENCE: 10448-189001  
; CURRENT APPLICATION NUMBER: US/10/162,435  
; CURRENT FILING DATE: 2002-06-04  
; PRIOR APPLICATION NUMBER: US 09/836,499  
; PRIOR FILING DATE: 2001-04-17  
; PRIOR APPLICATION NUMBER: PCT/US01/12420  
; PRIOR FILING DATE: 2001-04-17  
; PRIOR APPLICATION NUMBER: US 60/197,507  
; PRIOR FILING DATE: 2000-04-18

```
/ PRIOR APPLICATION NUMBER: US 09/891,008
/ PRIOR FILING DATE: 2001-06-25
/ PRIOR APPLICATION NUMBER: PCT/US01/19963
/ PRIOR FILING DATE: 2001-06-25
/ PRIOR APPLICATION NUMBER: US 60/214,220
/ PRIOR FILING DATE: 2000-06-23
/ PRIOR APPLICATION NUMBER: US 09/860,868
/ PRIOR FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: PCT/US01/16013
/ PRIOR FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: US 60/205,674
/ PRIOR FILING DATE: 2000-05-19
/ PRIOR APPLICATION NUMBER: US 09/886,429
/ PRIOR FILING DATE: 2001-06-21
/ PRIOR APPLICATION NUMBER: PCT/US01/20055
/ PRIOR FILING DATE: 2001-06-21
/ PRIOR APPLICATION NUMBER: US 60/213,963
/ PRIOR FILING DATE: 2000-06-23
/ PRIOR APPLICATION NUMBER: US 10/041,406
/ PRIOR FILING DATE: 2002-01-08
/ PRIOR APPLICATION NUMBER: PCT/US02/00275
/ PRIOR FILING DATE: 2002-01-08
/ PRIOR APPLICATION NUMBER: US 60/260,286
/ PRIOR FILING DATE: 2001-01-08
/ PRIOR APPLICATION NUMBER: US 09/934,268
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: PCT/US01/41811
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: US 60/226,612
/ PRIOR FILING DATE: 2000-08-21
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 35
/ LENGTH: 3184
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (168)...(977)
US-10-162-435-35

Alignment Scores:
Pred. No.: 2,75e-131 Length: 3184
Score: 1215.50 Matches: 227
Percent Similarity: 96.60% Conservative: 0
Best Local Similarity: 96.60% Mismatches: 6
Query Match: 94.74% Indels: 2
DB: 14 Gaps: 1

US-09-972-970-4 (1-233) x US-10-162-435-35 (1-3184)

Qy 1 MetProGlyLysHisGlnPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
Db 168 ATGCCCGGCAAGCAGCAGCATTTCCAGGACCTTCAGGCTCGGCTCGGGAATACTTC 227
Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db 228 CTGTTTGGCTTCAACATTGCTCTTCGGGTGCTGGAGCCCTGCTTCCTGCTATGGCCTC 287
Qy 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db 288 TGGGCTCGGGGAGAGGGCGTTCTCTCGAACATCTCAGCGCTCAGAGATCTGGGAGGC 347
Qy 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80
Db 348 CTTGACCCCGTGTGGCTGTTGTGTAGTTGGAGCGTCATGTCGCTCGGCTTTCCT 407
Qy 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
Db 408 GGCTGCATTTGGGGCCCTCCGGAGAACACCTTCCTGCTCAAGTTTTCCTCGTGTCTTC 467
Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
Db 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120

468 GGTCTCATCTTCTTCTGGAGCTGGCAACAGGATCCTGGCCCTTTGTCTTCAAGGACTGG 527
121 IleArgAspGlnLeuAsnPhePheIleAsnAsnAsnValLysAlaTyrArgAspAspIle 140
528 ATTGAGACCGAGCTCAACCTTTCATCAACAACAACAGTCAAGGCCCTACCGGAGCAGCAT 587
141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
588 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATATCTGCTTCTGCTGGAGCCCGAGGC 647
161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
648 CCAATGACTGGAACTCAATATCTACTTCAACTGCACTGACTTGAACCCCGGAG 707
181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199
708 CGCTCGGGGTGCTCTTCTCTGCTGCGTCAAGGACCTTCGCGAGGATGCTCTCAACACC 767
199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
768 CAGTGTGGCTACGACGTCGGGCTCAAACTGGAGCTGGAGCAGCAGGAGGCTTCATCCAC 827
219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
828 AAAGGCTGCTGGCGCAGTTTGAGAAAGTGCTGCAGCAACACC 870

RESULT 7
US-10-103-196-4
; Sequence 4, Application US/10103196
; Publication No. US20030050466A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT008P1
; CURRENT APPLICATION NUMBER: US/10/103,196
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/707,936
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/13504
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/178,770
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/149,447
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/138,573
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/137,797
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/135,122
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-103-196-4

Alignment Scores:
Pred. No.: 1.35e-131 Length: 1126
Score: 1212.50 Matches: 226
Percent Similarity: 96.60% Conservative: 1
Best Local Similarity: 96.17% Mismatches: 6
Query Match: 94.51% Indels: 2
DB: 14 Gaps: 1

US-09-972-970-4 (1-233) x US-10-103-196-4 (1-1126)

Qy 1 MetProGlyLysHisGlnPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
Db 157 ATGCCCGGCAAGCAGCAGCATTTCCAGGACCTTCAGGCTCGGCTCGGGAATACTTC 216
Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
```

```
Db 217 CTGTTTGGCTTCAACATTGCTTCTGGGTCTGGAGCCCTGTTCTCTGGCTATCGGCCCTC 276
Qy 41 TtpAlaTripglylulysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db 277 TGGGCTTGGGTGAGAAAGGCGTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 336
Qy 61 LeuAspProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80
Db 337 CTTCACCCCGTGGGCTGTTTGGTAGTGGAGGCGTCATGTCGGTGTGGGCTTTGCT 396
Qy 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePheSerValPheLeu 100
Db 397 GGCTGATTGGGCGCTCCGGGAGAACACCTTCTGCTCAAGTTTTCTCCGTTCTCCTC 456
Qy 101 GlyLeuIlePhePheLeuLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
Db 457 GGTCCTCATCTCTCTCGAGCTGGCAACAGGAGTCTGGCCCTTGTCTTCAAGGACTCG 516
Qy 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValIleValAlaTyrrArgAspAlle 140
Db 517 ATTCGAGACAGCTCAACCTCTTCATCAACAACAGCTCAAGCCCTACCGGAGCAGATT 576
Qy 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrrTrpSerCysCysGlyAlaArgGly 160
Db 577 GACCTCAGAACCTCATTGACTTTGCTCAGGAATACTGGTCTGTGTCGAGCCCGAGGC 636
Qy 161 ProAsnAspTrpAsnLeuAsnIleTyrrPheAsnCysThrAspLeuAsnProSerArgGlu 180
Db 637 CCCAATGACTGGAACTCAATATCTACTTCACTCAACAACAGCTCAAGTGAACCCAGCGGAG 696
Qy 181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199
Db 697 CGCTGCGGGTGGCCCTTCTCTGCTCGTCAGGAGCCCTGGGAGAGATGCTCTCAACACC 756
Qy 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrrThrPr 219
Db 757 CAGTGTGGGTACGACGTCGGGCTCAAACTGGAGCTGGAGCAGCAGGCGCTTCATCCACACC 816
Qy 219 olysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
Db 817 AAAGGCTGCTGGGCGAGTTTGAGAAATGGCTGCAGGACAACC 859
```

## RESULT 8

```
US-09-925-299-205
; Sequence 205, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 205
; LENGTH: 1655
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1548)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1559)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1564)
; OTHER INFORMATION: n equals a,t,g, or c
```

```
; NAME/KEY: misc_feature
; LOCATION: (1623)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1643)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-205

Alignment Scores:
Pred. No.: 1.09e-101 Length: 1655
Score: 958.50 Matches: 180
Percent Similarity: 84.19% Conservative: 17
Best Local Similarity: 76.92% Mismatches: 33
Query Match: 74.71% Indels: 4
DB: Gaps: 2
```

US-09-972-970-4 (1-233) x US-09-925-299-205 (1-1655)

```
Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrrPhe 20
Db 277 ATGTCGGGAAG-----CACTACAAAGGTCCTGAAGTCAGTTGTTGCATCAAAATCTTC 330
Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db 331 ATATTGGCTTCATGTCATATTTGGTTTTGGGAATAACATTTCTTGGAAATTGAGCTG 390
Qy 41 TtpAlaTripglylulysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db 391 TGGGATGGAATGAAAGAGGAGTCTGTCCAACAATCTCTTCCATCACCGCATCTCGGCGGC 450
Qy 61 LeuAspProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80
Db 451 TTTGACCCAGTTTGGCTCTCTCTGTGGTGGAGGAGTGATGTTCAATTTGGGATTTGCA 510
Qy 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePheSerValPheLeu 100
Db 511 GGTCGATTGGAGCCCTACGAGAAACACTTTCCTTCTCAAGTTTTTTCTGTGTTCTCTG 570
Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
Db 571 GGAATTAATTTCTCTCGAGCTCACTGCCGAGTTCTAGCATTTGTTTTCAAAGACTGG 630
Qy 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValIleAlaTyrrArgAspAlle 140
Db 631 ATCAAGACCCAGCTGTATTTCTTTATAACAACAACATCAGAGCATATCGGATGCAATT 690
Qy 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrrTrpSerCysCysGlyAlaArgGly 160
Db 691 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGTGTGGGCTTTTGA 750
Qy 161 ProAsnAspTrpAsnLeuAsnIleTyrrPheAsnCysThrAspLeuAsnProSerArgGlu 180
Db 751 GCTGATGATTGGAACTAAATATTATTACATTCAATTCACAGATTCCAATCCAAGTCGAGAG 810
Qy 181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199
Db 811 CGATGTGCGTTCCATTCTCTGCTGCACATAAAGATCCCGCAAGAGATGTCATCAACACT 870
Qy 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyrrThrPr 219
Db 871 CAGTGTGCTATGATGCCAGGCAAAAACAGAGTTGACCAGCATTTGATTAATCTACACG 930
Qy 219 olysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
Db 931 AAAGGCTGTGCCCCAGTTTGGAGAAAGTGTGTCAGGACA 970
```

## RESULT 9

```
US-09-925-299-205
; Sequence 205, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
```



FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 205  
LENGTH: 1655  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1548)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1559)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1564)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1623)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1643)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-299-205

Alignment Scores:  
Pred. No.: 1,09e-101 Length: 1655  
Score: 958.50 Matches: 180  
Percent Similarity: 84.19% Conservative: 17  
Best Local Similarity: 76.92% Mismatches: 33  
Query Match: 74.71% Indels: 4  
DB: 11 Gaps: 2

US-09-972-970-4 (1-233) x US-09-925-299-205 (1-1655)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
Db 277 ATGTCCGGGAAG-----CACTACAGGGTCTCTGAAGTCAGTTGTCATCAAAATCTTC 330  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 331 ATATTTGGCTTCAATGTCATATTTTGGTTTGGGAATAACATTTCTTGGAAATTGGACTG 390  
QY 41 TrpAlaTrpGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 391 TGGGCATGGAATCAAAAAGGAGTTCTGTCCACATCTTCCATCACCATCTCCGGCGC 450  
QY 61 LeuAppProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80  
Db 451 TTTGACCCAGTTGGCTCTCTCTGTGTGGGAGGAGTCATGTTCAATTTGGGATTGCA 510  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLysPhePheSerValPheLeu 100  
Db 511 GGGTGATTTGGAGCGCTACGGAAACATCTTCCTCTCAAGTTTTTCTGTGTCTCG 570  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 571 GGAATATTCTTCTCGAGCTCACTGCGGAGTTCTAGCATTTGTTTCAAAAGACTGG 630  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgPheAspIle 140  
Db 631 ATCAAGACCACTGATTTCTTTTATAAACAACATCAGAGCATATCGGGATGACATT 690  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaIleGlyLeu 160  
Db 691 GATTTGCCAAACCTCATAGACTTCCCGAGGAATATTGGCAGTGTCTGTGGGCTTTTGA 750

QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 751 GCTGATGATTGGAACTTAATAATTTACTTCAATTGCACAGATTCCTCAATGCAAGTCGAGAG 810  
QY 181 AtgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199  
Db 811 CGATGTGGGTTCATTTCTCTGCTGACATAAAGATCCCGCAGAGATGTCATCAACACT 870  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219  
Db 871 CAGTGTGGCTATGATGCCAGCAAAACCAGAAAGTTGACCAGCAGATTGTAATCTACACG 930  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
Db 931 AAAGCTGTGTGCCCCAGTTTGAGAAAGTGTTCAGGACA 970

## RESULT 10

US-10-103-196-5  
Sequence 5, Application US/10103196  
Publication No. US20030050466A1  
GENERAL INFORMATION:  
APPLICANT: Ni et al.

TITLE OF INVENTION: TMA5F Polynucleotides, Polypeptides, and Antibodies  
FILE REFERENCE: PT008P1  
CURRENT APPLICATION NUMBER: US/10/103,196  
CURRENT FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: 09/707,936  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: PCT/US00/13504  
PRIOR FILING DATE: 2000-05-18  
PRIOR APPLICATION NUMBER: 60/178,770  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: 60/149,447  
PRIOR FILING DATE: 1999-08-18  
PRIOR APPLICATION NUMBER: 60/138,573  
PRIOR FILING DATE: 1999-06-11  
PRIOR APPLICATION NUMBER: 60/137,797  
PRIOR FILING DATE: 1999-06-03  
PRIOR APPLICATION NUMBER: 60/135,122  
PRIOR FILING DATE: 1999-05-19  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 1174

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (2)  
OTHER INFORMATION: n equals a,t,g, or c  
US-10-103-196-5

Alignment Scores:  
Pred. No.: 3.09e-97 Length: 1174  
Score: 918.50 Matches: 179  
Percent Similarity: 83.40% Conservative: 17  
Best Local Similarity: 76.17% Mismatches: 34  
Query Match: 71.59% Indels: 6  
DB: 14 Gaps: 2

US-09-972-970-4 (1-233) x US-10-103-196-5 (1-1174)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
Db 93 ATGTCGGGAAG-----CACTACAGGGTCTCTGAAGTCAGTTGTCATCAAAATCTTC 146  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 147 ATATTTGGCTTCAATGTCATATTTGTTTGGGAATAACATTTCTTGGAAATTGGACTG 206

QY 41 TTPAlaTTPGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 207 TGGGATGGAAATGAAAGAGTGTCTGTCACCATCTTCCATCCAGCATCTCGGGCGC 266  
QY 61 LeuAspProValTTPLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 267 TTTGACCCAGCTTTGGCTCTTCTTGTGTGGAGGAGTGATGTTTCAATTTGGGATTTGCA 326  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPheSerValPheLeu 100  
Db 327 GGGTGCATTTGGAGCGTA-CGGGAAACACATTTCTTCTCAAGTTTTTTTCTGTGTTCTCG 385  
QY 101 GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120  
Db 386 GGAATTAATTTCTTCTGTGAGCTCACTGCGGAGTTCTAGCATTTGTTTCAAGACTG 445  
QY 120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAl 140  
Db 446 GATCAAGACCCAGCTGTATTCTTTTATAAACAACATCAGAGCATATCGGATGCAT 505  
QY 140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgG 160  
Db 506 TGATTTGCAAAACCTCATAGACTTCAACCAGGAATATTGGCAGTGTGTTGGGCTTTGG 565  
QY 160 YProAsnAspTTPAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgG 180  
Db 566 AGCTGATGATTGAACCTAAATATTACTTCAATTCACAGATTCCAATGCCAAGTCGAGA 625  
QY 180 uArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThP 199  
Db 626 GCGATGTGGGTTCCATTCTCTGCTGCCTAAAGATCCGCGAGAGATGTCATCAACAC 685  
QY 199 roSerValAlaMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyThP 219  
Db 686 TCAGTGTGGCTATGTGCCGCGGCAAAACAGAAAGTTGACACAGATTGTAATCTACAC 745  
QY 219 roLysAlaValTTPAlaSerLeuArgSerGlyCysArgThr 232  
Db 746 GAAAGGCTGTGCCCCAGTTTGAGAAAGTGGTTGCAGGACA 786

## RESULT 11

US-10-103-196-10  
; Sequence 10, Application US/10103196  
; Publication No. US20030050466A1  
; GENERAL INFORMATION:  
; APPLICANT: NI et al.  
; TITLE OF INVENTION: TM4SF Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PT008P1  
; CURRENT APPLICATION NUMBER: US/10/103,196  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/707,936  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: PCT/US00/13504  
; PRIOR FILING DATE: 2000-05-18  
; PRIOR APPLICATION NUMBER: 60/178,770  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/149,447  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/138,573  
; PRIOR FILING DATE: 1999-06-11  
; PRIOR APPLICATION NUMBER: 60/137,797  
; PRIOR FILING DATE: 1999-06-03  
; PRIOR APPLICATION NUMBER: 60/135,122  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 1178  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-103-196-10

## Alignment Scores:

Pred. No.: 3,37e-94 Length: 1178  
Score: 892.50 Matches: 179  
Percent Similarity: 82.70% Conservative: 17  
Best Local Similarity: 75.53% Mismatches: 34  
Query Match: 69.56% Indels: 8  
DB: 14 Gaps: 2  
US-09-972-970-4 (1-233) x US-10-103-196-10 (1-1178)  
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValValCysGlyLysTyrPhe 20  
Db 91 ATGTCCGGAG-----CACTACAAGGCTCTGTAAGTCAGTTGTGCATCAATACTTC 144  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 145 ATATTTGGCTTCAATGTCATATTTTGGTTTTTGGGAATAACATTTCTTGGAAATTGGACTG 204  
QY 41 TTPAlaTTPGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 205 TGGGATGGAATGAAAGAGTTCTTCCATCATCTTCCATCACCAGATCTCGGGCGC 264  
QY 61 LeuAspProValTTPLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 265 TTTGACCCAGTTTGGCTCTTCTTGTGTGGAGGAGTGATGTTTCAATTTGGGATTTGCA 324  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSer-ValPheLe 100  
Db 325 GGGTGCATTTGGAGCGTA-CGGGAAACACATTTCTTCTCAAGTTTTTTTCTCGTTCCT 383  
QY 100 u-GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAsp 119  
Db 384 CGGGAATTAATTTTCTTCTGAGCTCACTGCGGAGTTCTAGCATTTGTTTCAAGAC 443  
QY 120 TTPAlaArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAsp 139  
Db 444 TGGATCAAGACCCAGCTGTATTCTTTATAAACAACATCAGACCATATCGGATGAC 503  
QY 140 IleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArg 159  
Db 504 ATTGATTTGCAAAACCTCATAGACTTCAACCAGGAATATTGGCAGTGTCTGTTGGGCTTT 563  
QY 160 GlyProAsnAspTTPAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArg 179  
Db 564 GGAGCTGATGATTGGAACCTAAATATTACTTCAATTCACAGATTCCTCAATGCAAGTCGA 623  
QY 180 GluArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerTh 198  
Db 624 GAGCGATGTGGGTTCCATTCTCTGCTGCATTAAGATCCCGCAGAAAGATGTCATCAAC 683  
QY 198 rProSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyTh 218  
Db 684 ACTCAGTGTGGCTATGATGCCAGGCAAAACAGAAAGTTGACAGCAGATTGTAATCTAC 743  
QY 218 rProLysAlaValTTPAlaSerLeuArgSerGlyCysArgThr 232  
Db 744 ACGAAAGGCTGTGTGCCCGAGTTTGAGAAAGTGGTTGCAGGACA 786

## RESULT 12

US-09-729-674-169  
; Sequence 169, Application US/09729674  
; Patent No. US2001003935A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LaVallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steindinger II, Robert J.  
; APPLICANT: Spaulding, Vikki

APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729,674  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539,330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 169  
LENGTH: 1110  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-729-674-169

Alignment Scores:  
Pred. No.: 2,99e-84 Length: 1110  
Score: 807.00 Matches: 163  
Percent Similarity: 76.50% Conservative: 16  
Best Local Similarity: 69.66% Mismatches: 30  
Query Match: 62.90% Indels: 26  
DB: 9 Gaps: 3

US-09-972-970-4 (1-233) x US-09-729-674-169 (1-1110)

QY 1 MetProGlyLysHisGlnHisGlnGluProGluValGlyCysCysGlyLysTyrPhe 20  
DB 176 ATGTCGCGGAAG-----CACTACAGGGTCTCGAAGTCAGTTGTGCAATCAATCTTC 229

QY 21 LeuPheGlyPheAsnIleValPheTyrValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
DB 230 ATATTGGCTTCAATGCATATTTTGGTTTGGGAATAACATTCTTGGAAATTGGACTG 289

QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGly 60  
DB 290 TGGGCATGGAATGAAAGAGGAGTCTGTCCACATCTCTTCCATCACCAGATCTCGCGGC 349

QY 61 LeuAspProValTrpPheValValGlyValMetSerValLeuGlyPheAla 80  
DB 350 TTTGACCCAGTTTGGCTCTCTCTGTGGGAGGAGTGATGTTTCATTTGGGATTGCA 409

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
DB 410 GGGTGCAATTGGCGCTACGGGAAACACTTCTCTCTCAAGTTTCTTGTGTCTCTG 469

QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
DB 470 GGAATTATTTCTTCTCGAGCTCACTCGCGAGTCTTAGCATTTGTTTCAAAGACTGG 529

QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspIle 140  
DB 530 ATCAAGACCACTGATTTCTTTATAAACACACATCAGAGCATATCGGGATGACATT 589

QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160  
DB 590 GATTGCAAAACCTCATAGACTTCACCCAGGAATATAT-TCC----- 630

QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
DB 631 -----AATGCAAGTCGAGAG 645

QY 181 ArgCysGlyValProPheSerCysValArgAspProAla---MetSerSerThrPr 199  
DB 646 CGATGTGGCGTTCATTTCTCTGTGCTAAAGATCCCGCAGAGATGTCATCAACACT 705

QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyrThrPr 219  
DB 706 CAGTGTGGCTATGATGCCAGGCAAAACACAGAGTTGACCAGCAGATGTTAATCTACAG 765

QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232

DB 766 AAAGGCTGTGTGCCCCAGATTTCAGAGAGTGTGTTGCAGGACA 805

RESULT 13  
US-09-822-846-126  
Sequence 126, Application US/09822846  
Publication No. US20030027139A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: LaValle, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Agostino, Michael J.  
APPLICANT: Steinger II, Robert J.  
APPLICANT: Bowman, Michael R.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulukota, Kamalakara  
APPLICANT: Graham, James R.  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6400  
CURRENT APPLICATION NUMBER: US/09/822,846  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195,605  
PRIOR FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 629  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 126  
LENGTH: 1988  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-846-126

Alignment Scores:  
Pred. No.: 3.07e-83 Length: 1988  
Score: 801.50 Matches: 153  
Percent Similarity: 85.94% Conservative: 12  
Best Local Similarity: 79.69% Mismatches: 25  
Query Match: 62.47% Indels: 2  
DB: 11 Gaps: 1

US-09-972-970-4 (1-233) x US-09-822-846-126 (1-1988)

QY 43 TrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAsp 62  
DB 2 TGGAAATGAAAGAGGATCTGTCCAAACATCTCTTCCATCACCAGATCTCGCGGCTTTCGAC 61

QY 63 ProValTrpLeuPheValValGlyValMetSerValLeuGlyPheAlaGlyCys 82  
DB 62 CCAGTTTGGCTCTCTCTGTGGGAGGAGTGTGTTTCTTTGGGATTTTCAGGGTGC 121

QY 83 IleGlyAlaLeuArgGluAsnThrPheLeuLysPhePheSerValPheLeuGlyLeu 102  
DB 122 ATTGAGCGCTACCGGAAACACCTTCTTCTCAAGTTTCTTCTGTTTCTCTGGGAATT 181

QY 103 IlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheAspTrpIleArg 122  
DB 182 ATTTTCTTCTTGGAGCTCACTGCGGAGTCTAGCATTTTGTTCAAAGACTGGATCAAA 241

QY 123 AspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIleAspLeu 142  
DB 242 GACCAGCTGATTTCTTTTATAAACACACATCAGAGCATATCGGATGACATTTGATTG 301

QY 143 GlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGlyProAsn 162  
DB 143 GlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGlyProAsn 162

Db 302 CAAACCTCATAGACTTACCCAGGATATTGGCAGTGTGGGGCTTTTGGAGCTGAT 361  
Qy 163 AspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGluArgCys 182  
Db 362 GATTGGAACCTAAATATTACTTCAATTGCACAGATTCCAATGCAAGTCGAGAGCGATGT 421  
Qy 183 GlyValProPheSerCysCysValArgAspProAla---MetSerSerThrProSerVa 201  
Db 422 GCGTTTCCATTTCCTGCTGCTACTAAAGATCCCGCAGAGATGTCTATCAACATCAGTGT 481  
Qy 201 llaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAl 221  
Db 482 GCGTATGATGCCAGGCAAAACAGAAAGTTGACCAGCAGATTGTATCTACACGAAAGCG 541  
Qy 221 aValTrpAlaSerLeuArgSerGlyCysArgThr 232  
Db 542 TGTGTCCCCAGTTTGAGAAGTGTGTTGCAGGACA 575

RESULT 14  
US-09-875-440-1  
; Sequence 1, Application US/09875440  
; Patent No. US20020156035A1  
; GENERAL INFORMATION:  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Jefferson, Anne B.  
; APPLICANT: Winter, Jill A.  
; APPLICANT: Randazzo, Filippo  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING  
; FILE OF INVENTION: NEOPLASTIC DISEASE USING NET-4 MODULATORS  
; FILE REFERENCE: PP-01701.002/200130.522  
; CURRENT APPLICATION NUMBER: US/09/875,440  
; CURRENT FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 864  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: NET-4 oligonucleotide used in cell proliferation  
US-09-875-440-1

Alignment Scores:  
Pred. No.: 2e-82 Length: 864  
Score: 790.00 Matches: 139  
Percent Similarity: 88.51% Conservative: 15  
Best Local Similarity: 79.89% Mismatches: 18  
Query Match: 61.57% Indels: 2  
DB: 10 Gaps: 1

US-09-972-970-4 (1-233) x US-09-875-440-1 (1-864)  
Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
Db 348 ATGTCCGGGAAG-----CACTACAGAGGTCTCGAAGTCAGTTGTTGCATCAAACTTC 401  
Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 402 ATATTGGCTTCAATGCTCATATTTTGGTTTGGGAATAACATTTCTTGGAAATTGGACTG 461  
Qy 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 462 TGGGATGGAATGAAAGAGAGTTCTGTCCACATCTCTTCATCACCAGTCTCGGCGGC 521  
Qy 61 LeuAppProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 522 TTTGACCCAGTTTGGCTCTCTCTGTTGGTGGGAGGAGTGTATCTATTTTGGATTGCA 581  
Qy 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 582 GGGTGCAITGGAGCGCTACGGGAACACTTTTCTCTCAAGTTTTTTCTGTGTCTCTG 641

Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 642 GGAATTAATTTTCTTCTGAGCTACTGCCGAGTTCTAGCATTTGTCTTCAAGACTGG 701  
Qy 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValIleAlaTyrArgAspAspIle 140  
Db 702 ATCAAGACCACTGTATTTCTTTATTAACAACAACATCAGAGCATATCGGATGACATT 761  
Qy 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160  
Db 762 GATTTCGAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGTCTGGGCTTTTGA 821  
Qy 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174  
Db 822 GCTGATGATTGGAACCTAAATATTACTTTCAATTGCACAGAT 863

RESULT 15  
US-09-905-674-1  
; Sequence 1, Application US/09905674  
; Publication No. US20030039647A1  
; GENERAL INFORMATION:  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Garcia, Pablo  
; TITLE OF INVENTION: TETRASPAN PROTEIN AND USES THEREOF  
; FILE REFERENCE: PP-01700.002/200130.521  
; CURRENT APPLICATION NUMBER: US/09/905,674  
; CURRENT FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1388  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1285\_1377  
; OTHER INFORMATION: n = A,T,C or G  
; NAME/KEY: misc feature  
; LOCATION: 1285\_1377  
; OTHER INFORMATION: n = A,T,C or G  
US-09-905-674-1

Alignment Scores:  
Pred. No.: 2.67e-66 Length: 1388  
Score: 654.50 Matches: 125  
Percent Similarity: 70.13% Conservative: 37  
Best Local Similarity: 54.11% Mismatches: 67  
Query Match: 51.01% Indels: 2  
DB: 11 Gaps: 1

US-09-972-970-4 (1-233) x US-09-905-674-1 (1-1388)  
Qy 5 HisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPheLeuPheGlyPhe 24  
Db 73 TATTATAGTACTCTTAACGCCAAGTCAAGTCTGCTGGTCAAGTACTCTCTTTTCAGCTAC 132  
Qy 25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44  
Db 133 AACATCATCTTCTGCTGGCTGGAGTGTCTTCTTCTTGGAGTCGGCTGTGGCATGGAGC 192  
Qy 45 GlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProVal 64  
Db 193 GAAAGGGTGTGCTGCTCGACCTCACCAAGAGTACCCGAGTCATCGGAATCGACCTGTG 252  
Qy 65 TrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84  
Db 253 GTGCTGGTCTCGATGCTGGGGGTGGTGTATGTTTACCCTTGGGGTTCGCGGCTCGTGGG 312  
Qy 85 AlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeuIlePhe 104  
Db 313 GCTCTGCGGAGAAATATCTGCTTGTCTCAACTTTTCTGTGGCACCATCGTCTCATCTTC 372  
Qy 105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArgAspGln 124

[illegible]

Search completed: November 21, 2003, 17:20:02  
Job time : 340 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2003, 15:29:28 ; Search time 66 Seconds

(without alignment)  
1558.216 Million cell updates/sec

Title: US-09-972-970-4

Perfect score: 1283

Sequence: 1 MPGKHQHPQEPVEVCGCKGYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09972970/runat\_21112003\_125336\_27250/app\_query.fasta\_1.391  
-DB=Issued Patents NA -QEMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=BITS -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09972970@cgn 1.1.56@runat\_21112003\_125336\_27250 -NCPU=6 -ICPU=3  
-NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	476.5	37.1	1932	4	US-09-482-273-41
2	266	20.7	399	3	US-09-188-930-62
3	266	20.7	399	3	US-09-188-930-243
4	266	20.7	399	4	US-09-312-283C-62
5	266	20.7	399	4	US-09-312-283C-243
6	250.5	19.5	827	3	US-09-333-599-5
7	250.5	19.5	827	4	US-09-499-781-5
8	250.5	19.5	870	3	US-09-333-599-1
9	250.5	19.5	870	4	US-09-499-781-1
10	249.5	19.4	1344	3	US-08-705-771-8
11	248	19.3	1452	2	US-08-807-044-4
12	248	19.3	1452	5	PCT-US91-04986-1

13	248	19.3	1624	3	US-08-430-225A-19	Sequence 19, Appl
14	232	18.1	977	2	US-08-855-140-2	Sequence 2, Appl
15	232	18.1	977	4	US-09-016-434-938	Sequence 318, App
16	228	17.8	1151	2	US-08-807-044-2	Sequence 2, Appl
17	207.5	16.2	687	1	US-08-254-493-2	Sequence 2, Appl
18	207.5	16.2	687	1	US-08-408-222B-2	Sequence 2, Appl
19	207.5	16.2	1120	1	US-08-408-222B-3	Sequence 3, Appl
20	205.5	16.0	1120	1	US-08-254-493-3	Sequence 3, Appl
21	190	14.8	1001	3	US-08-705-771-6	Sequence 6, Appl
22	186	14.5	473	4	US-09-702-705-1565	Sequence 1565, Ap
23	186	14.5	473	4	US-09-736-457-1565	Sequence 1565, Ap
24	185.5	14.5	560	3	US-09-221-298-12	Sequence 12, Appl
25	183.5	14.3	933	3	US-08-808-148-2	Sequence 2, Appl
26	183.5	14.3	1289	3	US-09-020-956-111	Sequence 11, App
27	183.5	14.3	1289	3	US-09-030-607-111	Sequence 11, App
28	183.5	14.3	1289	4	US-09-439-313-111	Sequence 11, App
29	183.5	14.3	1289	4	US-09-352-616A-111	Sequence 11, App
30	183.5	14.3	1289	4	US-09-232-149A-111	Sequence 11, App
31	165	12.9	398	3	US-09-385-982-273	Sequence 273, App
C 32	155.5	12.1	801	3	US-09-020-956-16	Sequence 16, Appl
C 33	155.5	12.1	801	3	US-09-030-607-16	Sequence 16, Appl
C 34	155.5	12.1	801	4	US-09-439-313-16	Sequence 16, Appl
C 35	155.5	12.1	801	4	US-09-352-616A-16	Sequence 16, Appl
C 36	155.5	12.1	801	4	US-09-232-149A-16	Sequence 16, Appl
37	152	11.8	740	3	US-09-020-956-17	Sequence 17, Appl
38	152	11.8	740	3	US-09-030-607-17	Sequence 17, Appl
39	152	11.8	740	4	US-09-439-313-17	Sequence 17, Appl
40	152	11.8	740	4	US-09-352-616A-17	Sequence 17, Appl
41	152	11.8	740	4	US-09-232-149A-17	Sequence 17, Appl
42	134.5	10.5	1782	4	US-09-149-476-120	Sequence 120, App
43	118.5	9.2	1966	3	US-08-957-130-14	Sequence 14, Appl
C 44	117.5	9.2	751	3	US-09-020-956-12	Sequence 12, Appl
C 45	117.5	9.2	751	3	US-09-030-607-12	Sequence 12, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-482-273-41  
; Sequence 41, Application US/09482273  
; Patent No. 6534631  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 71 Human Secreted Proteins  
; FILE REFERENCE: P2030P1  
; CURRENT APPLICATION NUMBER: US/09/482,273  
; EARLIER FILING DATE: 2000-01-13  
; EARLIER APPLICATION NUMBER: PCT/US99/15849  
; EARLIER FILING DATE: 1999-07-14  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/092,922  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/092,956  
; EARLIER FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 267  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 1932  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (2)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1022)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-482-273-41

Alignment Scores:

Pred. No.: 2.75e-41 Length: 1932

```
Score: 476.50 Matches: 104
Percent Similarity: 57.61% Conservative: 36
Best Local Similarity: 42.80% Mismatches: 70
Query Match: 37.14% Indels: 33
DB: 4 Gaps: 4

US-09-972-970-4 (1-233) x US-09-482-273-41 (1-1932)

QY 3 GlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPheLeuPhe 22
Db 133 GGGAGGAGGTTCTCTTCCTGTCAGCCGCTGGTG-----AAATACCTGCTCTTC 180
QY 23 GlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAla 42
Db 181 TTCTTCAACATGCTCTTCGTTGATTTCCATGGTGTGGTGGCTGTGGTGTCTACGCT 240
QY 43 TrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyLeuAsp 62
Db 241 CGGCTAATGAAGCATGCAGAACGAGCCCTAGCTGCTG-----GCAGTGGAC 288
QY 63 ProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAlaGlyCys 82
Db 289 CTGCCATCTCTGCTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 348
QY 83 IleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeu 102
Db 349 ATTGGTCCCTCCGCGAGAACATCTGCTCTCTGCGAGAGCTTCTCCCTGCTGCTGCT 408
QY 103 IlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArg 122
Db 409 GTGTTCTCTGCTGAGTGGCGCTGGGATCCTGGGCTTCTGCTCTCTCAGAACAGGCTCG 468
QY 123 AspGlnLeuAsnPhePheIleAsnValLysAlaTyrArgAspIleLeu 142
Db 469 GGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 528
QY 143 GlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGlyProAsn 162
Db 529 CAGAACCTATTGATTTGGCGAGAAAGTTTAGCTGTGGAGGATTTCTCTACACAG 588
QY 163 AspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGluArgCys 182
Db 589 GACTGTCTCAGAACATGATTTCACTGCTCAGAACACCCAGTGCAGAGGCTGCTG 648
QY 183 GlyValProPheSerCysValArgAspPro----- 193
Db 649 TCTGTGCTTACTCTCTGTTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 708
QY 194 -----AlaMetSerSerThrProSerVa 201
Db 709 GGCCAAAGGTATGACGCTTTGACTACTTGGAAAGCTAGCAAAAGTCTATACCAATGCG 768
QY 201 AlaMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyrThrProLysAl 221
Db 769 TGTATGACAAAGT-----TGGTCAACTGGATACAGCAACCTATTCTTACTTGT 819
QY 221 aValTrp 223
Db 820 GTGTGG 826

RESULT 2
US-09-188-930-62
; Sequence 62, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1

US-09-972-970-4 (1-233) x US-09-188-930-62 (1-399)

QY 11 ProGluValGlyCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30
Db 7 CCGAAGTCAAGTGTGTGTCATCAATACCTTCTTTTGGCTTCAATGTCATATTTGGTTT 66
QY 31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSer 50
Db 67 TTGGAAATAACGTTTCTTGGAAATCGACTGTGGCGTGGAAATGAAAAAGGTGCTCTCC 126
QY 51 AsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValVal 70
Db 127 AACATCTGCTCCATCACCGACTCGTGGCTTTGAGCCAGTGTGGCTTTCTCTCAGTG 186
QY 71 -----GlyGlyValMetSerValLeuGly----- 78
Db 187 GCCAGCCGAGCGCTGAGCTCTGTCAATGACATCCAGGAGAGAAATCAGGTTAATGAGAGA 246
QY 79 -----PheAla 80
Db 247 CATTAATTAACACTCTCCCTCACCCACCGACCAACCGATGGTGGTCTTCTGTGATTTCT 306
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePhe-----SerVal 98
Db 307 GGAATACCTCTGGCTATG-----TTTATGTTTATTTCTTTTATTAATCGTTG 354
QY 99 PheLeuGlyLeuIlePhe 104
Db 355 TATTTTGGTCTTTTTTT 372

RESULT 3
US-09-188-930-243
; Sequence 243, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Mouse
US-09-188-930-243
Alignment Scores:
Pred. No.: 8.51e-20 Length: 399
```

Score: 266.00 Matches: 55  
Percent Similarity: 54.76% Conservative: 14  
Best Local Similarity: 43.65% Mismatches: 21  
Query Match: 20.73% Indels: 36  
DB: 4

US-09-972-970-4 (1-233) x US-09-188-930-243 (1-399)

QY 11 ProGluValGlyCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30  
DB 7 CCTGAAGTCAGTTGTCATCAATACTTTCATTTTGGCTTCAATGTCATATTTGGTTT 66

QY 31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGlyValLeuSer 50  
DB 67 TTGGGAATAACGTTTCTTGGAAATCGGACTGTGGCGTGGAAATGAAAAGGTGCTCTCC 126

QY 51 AsnIleSerAlaLeuThrAspLeuGlyLeuAspProValTrpLeuPheValVal 70  
DB 127 AACATCTCGTCCATCACCACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCTCTGAGTG 186

QY 71 -----GlyGlyValMetSerValLeuGly----- 78

DB 187 GCCAGCCGAGCTGAGCTCTGTCAATGACATCCAAAGGAGAAAATGAGTTAATGAGAGA 246

QY 79 -----PheAla 80

DB 247 CATTAATTAAACTCCCTCACCACCCGACCAAAACCAAGTGTGGTTCTTCTGATATCT 306

QY 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePhe-----SerVal 98  
DB 307 GGAATACTCTGGGCTATG-----TTTTATGTTTATTCTTTTAAATCGGTG 354

QY 99 PheLeuGlyLeuIlePhe 104  
DB 355 TATTTGGTCTTTT 372

## RESULT 4

US-09-312-283C-62  
; Sequence 62, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 62  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-312-283C-62

Alignment Scores:  
Pred. No.: 8,51e-20 Length: 399  
Score: 266.00 Matches: 55  
Percent Similarity: 54.76% Conservative: 14  
Best Local Similarity: 43.65% Mismatches: 21  
Query Match: 20.73% Indels: 36  
DB: 4

US-09-972-970-4 (1-233) x US-09-312-283C-62 (1-399)

QY 11 ProGluValGlyCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30  
DB 7 CCTGAAGTCAGTTGTCATCAATACTTTCATTTTGGCTTCAATGTCATATTTGGTTT 66

QY 31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGlyValLeuSer 50  
DB 67 TTGGGAATAACGTTTCTTGGAAATCGGACTGTGGCGTGGAAATGAAAAGGTGCTCTCC 126

QY 51 AsnIleSerAlaLeuThrAspLeuGlyLeuAspProValTrpLeuPheValVal 70  
DB 127 AACATCTCGTCCATCACCACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCTCTGAGTG 186

QY 71 -----GlyGlyValMetSerValLeuGly----- 78

DB 187 GCCAGCCGAGCTGAGCTCTGTCAATGACATCCAAAGGAGAAAATGAGTTAATGAGAGA 246

QY 79 -----PheAla 80

DB 247 CATTAATTAAACTCCCTCACCACCCGACCAAAACCAAGTGTGGTTCTTCTGATATCT 306

QY 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePhe-----SerVal 98  
DB 307 GGAATACTCTGGGCTATG-----TTTTATGTTTATTCTTTTAAATCGGTG 354

QY 99 PheLeuGlyLeuIlePhe 104  
DB 355 TATTTGGTCTTTT 372

## RESULT 5

US-09-312-283C-243  
; Sequence 243, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 243  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-312-283C-243

Alignment Scores:  
Pred. No.: 8,51e-20 Length: 399  
Score: 266.00 Matches: 55  
Percent Similarity: 54.76% Conservative: 14  
Best Local Similarity: 43.65% Mismatches: 21  
Query Match: 20.73% Indels: 36  
DB: 4

US-09-972-970-4 (1-233) x US-09-312-283C-243 (1-399)

QY 11 ProGluValGlyCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpVal 30  
DB 7 CCTGAAGTCAGTTGTCATCAATACTTTCATTTTGGCTTCAATGTCATATTTGGTTT 66

QY 31 LeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGlyGlyValLeuSer 50  
DB 67 TTGGGAATAACGTTTCTTGGAAATCGGACTGTGGCGTGGAAATGAAAAGGTGCTCTCC 126

QY 51 AsnIleSerAlaLeuThrAspLeuGlyLeuAspProValTrpLeuPheValVal 70  
DB 127 AACATCTCGTCCATCACCACCTCGGTGGCTTTGACCCAGTGTGGCTTTTCTCTGAGTG 186

QY 71 -----GlyGlyValMetSerValLeuGly----- 78



```
Db 187 GCCAGCCGAGCTGAGCTCTGTAATGACATCCAGGAGAAAATGAGGTTAATGAGAGA 246
QY 79 -----PheAla 80
Db 247 CATTAAATTAACACTCCCTCACCCCGCACCAAAACAGTGGGTCTTCTGATATCT 306
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePhe-----SerVal 98
Db 307 GGAATACTCGGGCTATG-----TTTATGTTTATTTCTTTTAAATCGGTG 354
QY 99 PheLeuGlyLeuIlePhe 104
Db 355 TATTTGGTCTTTTTTT 372

RESULT 6
US-09-333-599-5
; Sequence 5, Application US/09333599
; Patent No. 6245898
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/333,599
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (57)..(815)
US-09-333-599-5

Alignment Scores:
Pred. No.: 1,12e-17 Length: 827
Score: 250.50 Matches: 77
Percent Similarity: 44.18% Conservative: 33
Best Local Similarity: 30.92% Mismatches: 99
Query Match: 19.52% Indels: 40
DB: Gaps: 9

US-09-972-970-4 (1-233) x US-09-333-599-5 (1-827)
QY 8 PheGlnGluProGluValGlyCys-----CysGlyLysTyrPheLeuPheGlyPhe 24
Db 66 TTCAACGAGAAGAACACATGTGGCAGCGTTTGCTCAAGTACCTGCTGTTACTTAC 125
QY 25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44
Db 126 AATGTGCTTCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 185
QY 45 GluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyLeuAspProVal 64
Db 186 CTCACAGT-----GACTACATCAGCTGCTGGCTCAGGCACCTACCTGGCCACGCC 239
QY 65 TrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84
Db 240 TACATCTGCTGGTGGCGGCACATGCTGTCATGTCATGTCATGTCATGTCATGTC 299
QY 85 AlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeuIlePhe 104
Db 300 ACCTTCAAGAGCGTGGGAACCTGCTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
QY 105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArgAspGln 124
Db 360 CTGCTGGAGATCATCGTGGTATCTCGCTAGGCTAC-----TACCAGCAG 407
QY 125 LeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIleAsp----- 141
```

```
Db 408 CTGAACACGAGCTCAAGGAGAACCTGAAG-----GACACCATGACCAAGCGCTAC 459
QY 142 -----LeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSer 154
Db 459 CACCAGCCGGCCATGAGGCTGTGACCGGCTGTGGACCAAGCTGCAGCAGGAGTTCAC 518
QY 155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174
Db 519 TGTGTGGCAGCAACAACATCACAGGAGCTGCGCAGACAGTGGTGGATCCGC----- 569
QY 175 LeuAsnProSerArgGluArgCysGly-----ValProPheSerCysCys----- 189
Db 570 -----TCACAGGAGCGCGTGGCTGCTGCCAGACAGCTGCTGCAAGACGGTG 620
QY 190 -----ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204
Db 621 GTGGCTCTTTGTGGCAGCAGACCATGCTCCACATCTACAAGGTGGAGGCGGCTGC 680
QY 204 tSerGlySerAsnTrpSerTrpSerArgAlaProTyrThrProLysAlaValTrpAl 224
Db 681 ATCACCAGT---TGGAGACCTTCATCCAGGAGCACCTGAGGCTCATTTGGGCTGGGG 737
QY 224 aSerLeuArgSerGlyCysArgThr 232
Db 738 ATCGCATGCTGCTGTCAGGTCT 762

RESULT 7
US-09-499-781-5
; Sequence 5, Application US/09499781
; Patent No. 6498014
; GENERAL INFORMATION:
; APPLICANT: Testa, Jacqueline E.
; APPLICANT: Quigley, James P.
; APPLICANT: Seandel, Marco
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
; FILE REFERENCE: SUNY
; CURRENT APPLICATION NUMBER: US/09/499,781
; CURRENT FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 09/333,599
; PRIOR FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (57)..(815)
US-09-499-781-5

Alignment Scores:
Pred. No.: 1,12e-17 Length: 827
Score: 250.50 Matches: 77
Percent Similarity: 44.18% Conservative: 33
Best Local Similarity: 30.92% Mismatches: 99
Query Match: 19.52% Indels: 40
DB: Gaps: 9

US-09-972-970-4 (1-233) x US-09-499-781-5 (1-827)
QY 8 PheGlnGluProGluValGlyCys-----CysGlyLysTyrPheLeuPheGlyPhe 24
Db 66 TTCAACGAGAAGAACACATGTGGCAGCGTTTGCTCAAGTACCTGCTGTTACTTAC 125
QY 25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44
Db 126 AATGTGCTTCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 185
QY 45 GluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyLeuAspProVal 64
```

Db 186 CTCAGAGT-----GACTACATCAGCCTCTGGCTCAGGCACCTACCTGGCCACAGCC 239  
Qy 65 TrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84  
Db 240 TACATCTGTGTGGCGGCACTGTCTCATGTGACTGGGTCTTGGCTCTGTGGCC 299  
Qy 85 AlaLeuArgGluAsnThrPheLeuLeuLeuPheSerValPheLeuGlyLeuIlePhe 104  
Db 300 ACCTTCAAGAGGCGTCGGAACCTGCTGCTGCTTCTCATCTCTCTCTCATCATCTTT 359  
Qy 105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLeuLeuLeuLeuLeuLeu 124  
Db 360 CTGCTGGAGATCATCTGTGTATCTCTGGCTCAGCCTAC-----TACCAGCAG 407  
Qy 125 LeuAsnPhePheIleAsnAsnValLeuAlaTyzArgAspAspIleAsp----- 141  
Db 408 CTGAACACGAGGCTCAAGAGAACCTGAAG-----GACACCATGACCAAGCGCTAC 458  
Qy 142 -----LeuGlnAsnLeuIleAspPheAlaGlnGluTyTrpSer 154  
Db 459 CACCAGCGGGCCATCAGGCTGTGACGAGCGCTGTGGACGAGCTGCAGCAGGAGTTCAC 518  
Qy 155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyzPheAsnCysThrAsp 174  
Db 519 TGTCTGGGAGCAACAACTCAGAGGACTGGCGAGACAGTGGATCCGC----- 569  
Qy 175 LeuAsnProSerArgGluArgCysGly-----ValProPheSerCysCys 189  
Db 570 -----TCACAGAGCGCGGTGGCGGTGTGCTCCAGACAGCTGTGCAAGCGGTG 620  
Qy 190 -----ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204  
Db 621 GTGGCTCTTTTGGACAGCGAGACCATGCTCCCAACATCTACAAGTGGAGGCGGCTGC 680  
Qy 204 tSerGlySerAsnTrpSerTrpSerSerArgAlaProTyzThrProLyAlaValTrpAl 224  
Db 681 ATCACCAGT---TGGAGACCTTCATCCAGGACGACCTAGGCTCATTTGGGCTGTGGG 737  
Qy 224 aSerLeuArgSerGlyCysArgThr 232  
Db 738 ATCGCATTTGCCTGTGTGCAGGTCT 762

## RESULT 8

US-09-333-599-1  
; Sequence 1, Application US/09333599 ;  
; Patent No. 6245898  
; GENERAL INFORMATION:  
; APPLICANT: Testa, Jacqueline E.  
; APPLICANT: Quigley, James P.  
; APPLICANT: Seandel, Marco  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS  
; TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS  
; FILE REFERENCE: SUNY  
; CURRENT APPLICATION NUMBER: US/09/333,599  
; CURRENT FILING DATE: 1999-06-15  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 870  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-333-599-1

## Alignment Scores:

Pred. No.:	1,21e-17	Length:	870
Score:	250.50	Matches:	77
Percent Similarity:	44.18%	Conservative:	33
Best Local Similarity:	30.94%	Mismatches:	99
Query Match:	19.52%	Indels:	40
DB:	3	Gaps:	9

US-09-972-970-4 (1-233) x US-09-333-599-1 (1-870)

Qy 8 PheGlnGluProGluValGlyCys-----CysGlyLysTyzPheLeuPheGlyPhe 24  
Db 110 TTCACGAGAGAGAGACACATGTGGCAGCGTTTGGCTCAAGTACCTGCTGTTTACCTAC 169  
Qy 25 AsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeuTrpAlaTrpGly 44  
Db 170 AATTGTGCTTCTGGCTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 229  
Qy 45 GluLyGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAspProVal 64  
Db 230 CTCAGAGT-----GACTACATCAGCCTGTGGCTCAGGCACCTACCTGGCCACAGCC 283  
Qy 65 TrpLeuPheValValGlyValMetSerValLeuGlyPheAlaGlyCysIleGly 84  
Db 284 TACATCTGTGTGGCGGCACTGTCTCATGTGACTGGGTCTTGGCTGTGGCC 343  
Qy 85 AlaLeuArgGluAsnThrPheLeuLeuLeuPheSerValPheLeuGlyLeuIlePhe 104  
Db 344 ACCTTCAAGAGGCGTCGGAACCTGTCTGCTGCTGTCTCTCTCTCTCTCTCTCTCT 403  
Qy 105 PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLeuLeuLeuLeuLeuLeu 124  
Db 404 CTGCTGGAGATCATCTGCTGTATCTCTGCTTACGCTAC-----TACCAGCAG 451  
Qy 125 LeuAsnPhePheIleAsnAsnValLeuAlaTyzArgAspAspIleAsp----- 141  
Db 452 CTGAACACGAGGCTCAAGAGAACCTGAAG-----GACACCATGACCAAGCGCTAC 502  
Qy 142 -----LeuGlnAsnLeuIleAspPheAlaGlnGluTyTrpSer 154  
Db 503 CACCAGCGGGCCATCAGGCTGTGACGAGCGCTGTGGACGAGCTGCAGCAGGAGTTCAC 562  
Qy 155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyzPheAsnCysThrAsp 174  
Db 563 TGTCTGGGAGCAACAACTCAGAGGACTGGCGAGACAGTGGATCCGC----- 613  
Qy 175 LeuAsnProSerArgGluArgCysGly-----ValProPheSerCysCys 189  
Db 614 -----TCACAGAGCGCGGTGGCGGTGTGCTCCAGACAGCTGTGCAAGCGGTG 664  
Qy 190 -----ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204  
Db 665 GTGGCTCTTTTGGACAGCGAGACCATGCTCCCAACATCTACAAGTGGAGGCGGCTGC 724  
Qy 204 tSerGlySerAsnTrpSerTrpSerSerArgAlaProTyzThrProLyAlaValTrpAl 224  
Db 725 ATCACCAGT---TGGAGACCTTCATCCAGGACGACCTAGGCTCATTTGGGCTGTGGG 781  
Qy 224 aSerLeuArgSerGlyCysArgThr 232  
Db 782 ATCGCATTTGCCTGTGTGCAGGTCT 806

## RESULT 9

US-09-499-781-1  
; Sequence 1, Application US/09499781  
; Patent No. 6498014  
; GENERAL INFORMATION:  
; APPLICANT: Testa, Jacqueline E.  
; APPLICANT: Quigley, James P.  
; APPLICANT: Seandel, Marco  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS  
; TITLE OF INVENTION: ASSOCIATED WITH TUMOR METASTASIS  
; FILE REFERENCE: SUNY  
; CURRENT APPLICATION NUMBER: US/09/499,781  
; CURRENT FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 09/333,599  
; PRIOR FILING DATE: 1999-06-15  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 870  
; TYPE: DNA  
; ORGANISM: Homo sapiens

## US-09-499-781-1

## Alignment Scores:

Pred. No.: 1,21e-17 Length: 870  
Score: 250.50 Matches: 77  
Percent Similarity: 44.18% Conservative: 33  
Best Local Similarity: 30.92% Mismatches: 99  
Query Match: 19.52% Indels: 40  
DB: 4 Gaps: 9

## US-09-972-970-4 (1-233) x US-09-499-781-1 (1-870)

QY 8 PheGluProGluValGlyCys-----CysGlyValTyrPheLeuPheGlyPhe 24  
Db 110 TTCACGAGAGACACACATGTGGCAGCTTGGCTCAAGTACCTGCTGTACCTAC 169  
QY 25 AsnileValPheTrpValLeuGlyAlaLeuPheLeuAlaileGlyLeuTrpAlaTrpGly 44  
Db 170 AATTGCTGCTTCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 229  
QY 45 GluLeuGlyValLeuSerAsnileSerAlaLeuThrAspLeuGlyCysLeuAspProVal 64  
Db 230 CTCACAGAGT-----GACTACATCAGCCTGCTGGCTCAGGCACCTACCTGGCCACAGCC 283  
QY 65 TrpLeuPheValValGlyValMetSerValLeuGlyPheAlaGlyCysileGly 84  
Db 284 TACATCCTGGTGGCGGCACTGTCATGCTGAGTGGGTCTTGGGCTGGCTGGCC 343  
QY 85 AlaLeuArgGluAsnThrPheLeuLeuLysPheSerValPheLeuGlyLeuilePhe 104  
Db 344 ACCTTCAGGAGCGTCGGAACCTGCTGGCTGACTTCTCATCTCTCATCATCTTT 403  
QY 105 PheLeuGluLeuAlaThrGlyileLeuAlaPheValPheLysAspTrpIleArgAspGln 124  
Db 404 CTGCTGGAGATCATCGTGGTATCTCGCTACGCTAC-----TACCAGCAG 451  
QY 125 LeuAsnPhePheileAsnAsnValLysAlaTyrArgAspAspIleAsp----- 141  
Db 452 CTGACACGAGCTCAAGAGAACCTGAAG-----GACACATGACCAAGCGCTAC 502  
QY 142 -----LeuGlnAsnLeuileAspPheAlaGlnIleTrpSer 154  
Db 503 CACCAGCGCGCATGAGGCTGTGACGAGCGTGTGGACGAGTGCAGCAGGAGTTCCAC 562  
QY 155 CysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174  
Db 563 TCGTGTGGCAGCAACAATCAGCAGCTGGCGAGACAGTGTGATGGATCCGC----- 613  
QY 175 LeuAsnProSerArgGlyCysGly-----ValProPheSerCysCys----- 189  
Db 614 -----TCACAGGAGCGCGTGGCGTGTGGTCCAGACAGCTGTGCAAGACGGTG 664  
QY 190 -----ValArgAspProAlaMet-SerSerThrProSerValAlaMetMe 204  
Db 665 GTGGCTCTTTGTGGACGAGCAGACCATGCTCCCAACATCTACAAGGTGGAGGCGCTGC 724  
QY 204 tSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrProLysAlaValTrpAl 224  
Db 725 ATCACAAGT---TGGAGACCTTATCCAGGAGGACCTGAGGGTCATTGGGGCTGTGGG 781  
QY 224 aserLeuArgSerGlyCysArgThr 232  
Db 782 ATCGGCAATTGCTGTGTGACAGTCT 806

## RESULT 10

US-08-705-771-8

Sequence 8, Application US/08705771

Patent No. 6054289

GENERAL INFORMATION:

APPLICANT: Paul Moore, Reiner Gentz, Hongjin Ji,

APPLICANT: Jian Ni and Jing-Shan Hu

TITLE OF INVENTION: Human Genes, Sequences and

TITLE OF INVENTION: Expression Products

NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
FILING DATE: August 30, 1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: MULLINS, J.G.  
REGISTRATION NUMBER: 33,073  
REFERENCE/DOCKET NUMBER: 325800-346 (PFI96)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 973-994-1700  
TELEFAX: 973-994-1744  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1344 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-705-771-8

## Alignment Scores:

Pred. No.: 2,93e-17 Length: 1344  
Score: 249.50 Matches: 66  
Percent Similarity: 44.98% Conservative: 37  
Best Local Similarity: 28.82% Mismatches: 84  
Query Match: 19.45% Indels: 42  
DB: 3 Gaps: 9

## US-09-972-970-4 (1-233) x US-08-705-771-8 (1-1344)

QY 11 ProGluValGlyCysCysGly-----LysTyrPheLeu 21  
Db 123 CCAGAACTGAAGCGCTGGCGCATGGCGCGCTGCCCTCCAGCGCTCAAGTACCTCATG 182  
QY 22 PheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaileGlyLeuTrp 41  
Db 183 TTCGCTTCAACCTGCTCTTCTGGCTGGAGGCTGTGGCGTGTGGGCTGTGGCATCTGG 242  
QY 42 AlaTrpGlyGluLysGlyValLeuSerAsnileSerAlaLeuThrAspLeuGlyGlyLeu 61  
Db 243 CTGGCGGCACACAGGGGAGCTTGGCCACGCTGCTCT-----TCCTTC 287  
QY 62 AspProValTrp-----LeuPheValValGlyValMetSerValLeuGly 78  
Db 288 CCGTCCCTGGCTGGCTGGCAACCTGCTCATCACCGCGCGCTTGTTCATGGCCATCGGC 347  
QY 79 PheAlaGlyCysileGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerVal 98  
Db 348 TTCGTGGGCTGCTGGGGTGCATCAAGAGAACAGTGCCTCTCTCTCTCTCTCTCTCT 407  
QY 99 PheLeuGlyLeuilePhePheLeuLeuAlaThrGlyileLeuAlaPheValPheLys 118  
Db 408 CTGCTGTCTGTGGTGTCTCTCTGGAGGCGACCATCGCCATCTCTCTCTCTCTCTCTCT 467  
QY 119 AspTrpIleArgAspGlnLeuAsnPhePheileAsnAsnValLysAlaTyr----- 136  
Db 468 GACAAGATTGACAGGTATGCCACAGACCTGAAGAAAGGCTTGACACCTGTACGGCAGC 527  
QY 137 ArgAspIleAspLeuGlnAsnLeuileAspPheAlaGlnIleTyrTrpSerCysCys 156

```

Db 528 CAGGCAACGTGGGCTCACCAACGCCTGGAGCATCATCCAGCGACTTCGGCTGCTGT 587
Qy 157 GlyAlaArgGlyProAsnAspTrpAsnLeuAnlleTyrPheAsnCysThrAspLeuAn 176
Db 588 GCGTCTCCAACTACACTGACTGG---TTCGAGGTGTAC-----AAC 626
Qy 177 ProSerArgGluArgCysGlyValProPheSerCysCys----- 189
Db 627 GCCACGCG-----GTACCTGACTCTCTGCTTGGAGTTCAGTGAGAGCTGT 674
Qy 190 ---ValArgAspProAlaMet-----SerSerThrProSerValAlaMetMetSer 205
Db 675 GGGCTGCACGCCCGCGCACTGGTGAGGGCCGTGTACGAGAGGTGAAGGTGTGGCTTCA 734
Qy 206 Gly-SerAsnTrpSerSer 213
Db 735 GGAGNACTGTGGCTGTGGGCATCT 759

```

## RESULT 11

```

US-08-807-044-4
; Sequence 4, Application US/08807044
; Patent No 5863735
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

```

```

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0224 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

```

```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1452 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 180140
US-08-807-044-4

```

```

Alignment Scores:
Pred. No.: 4.75e-17 Length: 1452
Score: 248.00 Matches: 67
Percent Similarity: 48.39% Conservative: 38
Best Local Similarity: 30.88% Mismatches: 81
Query Match: 19.33% Indels: 31
DB: 2 Gaps: 9

```

```

US-09-972-970-4 (1-233) x US-08-807-044-4 (1-1452)
Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGly-----Lys 18
Db 44 GTGCTGGAAAGGCAAGAAATATACCGCATGGCGATGGCATGAGTAGCTTGAACATGCTGAAG 103
Qy 19 TyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIle 38
Db 104 TATGTCCTGTTTTCTTCACTTGTCTCTTTGGATCTGTGCTGTGCTGTGCTGCTTTGGGCTTT 163
Qy 39 GlyLeuTrp-----AlaTrpGlyGlyLysGlyValLeu-----SerAsnIleSerAlaLeu 55
Db 164 GGGATCTACCTGCTGATCCACAACAACATTCGGAGTGTCTTCCATAAACCTCCCTCCCTC 223
Qy 56 ThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValGlyGlyValMetSer 75
Db 224 ACG---CTGGGCAAT-----GTGTTTGTCTGCTGGGCTCTATTATCATG 265
Qy 76 ValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhe 95
Db 266 GTAGTTCCTTCCTGGGCTGCTGCTGCTTATCAAGGAAACAACTGCTGCTTATGTGCG 325
Qy 96 PheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPhe 115
Db 326 TTCTTCATCCTGCTGCTGATTATCTCTCTGCTGAGGTGACCTTGGCCATCTGCTCTTT 385
Qy 116 ValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAla 135
Db 386 GTATATGAACAGAAGCTGAATGAGTATGTGCTTAAGGTCTGACCCGACGATCCACCGT 445
Qy 136 TyrArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCys 155
Db 446 TACCACCTCAGACAATAGCACCAAGCGCGCTGGGACTCCATCCATCTGTCAGTGT 505
Qy 156 CysGlyAlaArgGlyProAsnAspTrpAsnLeuAnlleTyrPheAsnCysThrAspLeu 175
Db 506 TGTGCTATAATGGCAGCGAGTGATTGGACACGAGTGGCCACCAGCATCTTGC----- 556
Qy 176 AsnProSer-----ArgGluArg-CysGly-----ValPr 185
Db 557 ---CCCTCAGATCGAAAAGTGGAGGGTTGTATGCGAAAGCAAGACTGTGTTTCATTC 613
Qy 185 oPheSerCysCysValArgAspProAlaMetSerSerThrProSerVal 201
Db 614 AATTTCCTGT-----ATATCGGAATCATCACCATCTGTG 647

```

## RESULT 12

```

PCT-US91-04986-1
; Sequence 1, Application PC/TUS9104986
; GENERAL INFORMATION:

```

```

; APPLICANT: Seed, Brian
; APPLICANT: Allen, Janet
; APPLICANT: Aruffo, Alejandro
; APPLICANT: Camerini, David
; APPLICANT: Lauffer, Leander
; APPLICANT: Oquendo, Carmen
; APPLICANT: Simmons, David L.
; APPLICANT: Stamenkovic, Ivan
; APPLICANT: Stengelin, Siegfried
; APPLICANT: Amiot, Martine

```

```

; TITLE OF INVENTION: Rapid Immunoselection Cloning Method
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Greenlee & Associates
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```

Db      193 AAATACTTCTCTCTCTCAACTTGATCTTCTTATCTCTGGCGCAGTGATCTCTGGCG 252
QY      38 IleGlyLeuTrpAlaTrpGlyGluLys-----GlyValLeuSerAsnIleSer 53
Db      253 TTCGGGGTGTGGATCTCTGGCGCAGAGCAGTTTCATCTCTCTCTGCAAACTCTCTCC 312
QY      54 AlaLeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValGlyGlyVal 73
Db      313 AGCTCGCTTAGGATGGG-----GCCATGTCTTCTATCGCGCGTGGGGCAGTC 360
QY      74 MetSerValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeu 93
Db      361 ACTATGCTCATGGCTCTCTGGGCTGCATCGCGCGGTCAACGAGGTCCCTCGCTGCTG 420
QY      94 LysPhePheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeu 113
Db      421 GGGCTGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
QY      114 AlaPheValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnVal 133
Db      481 TTCCTACTTCAACATGGGCAAGCTGNAAGCAGGAGATGGCGGCATCGTGACTGACTCAT 540
QY      134 LysAlaTyr-----ArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGln 150
Db      541 CGAGACTACACAGCAGTCCGAGGAC---AGCTGAGGATGCCCTGGGACTACGTGCAG 597
QY      151 GluTyrTrpSerCysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPhe 170
Db      598 GCTCAGGTGAAGTGCTCGGC-----TGG-----GTCAGCTTCTAC 633
QY      171 AsnCysThrAsp-----LeuAsnProSerArgGluArgCysGlyValProPhe 186
Db      634 AACTGACAGACAAACGCTGAGCTCATGAAT-----CGCCTGAGGTCACTACCCCTGT 687
QY      187 SerCysCysValArg-----AspProAlaMetSer-----Ser 197
Db      688 TCTTGGGAAGTCAAGGGGAGAGGACACAGCCTTCTGTGAGGAAGGGCTTCTCGAG 747
QY      198 ThrProSerValAlaMetMetSerGlySer-----AsnTrpSerTrpSerArg 214
Db      748 GCCCCCGCAACAGGACCCAGAGTGGCAACACCCCTGAGGACTGG-----792
QY      215 AlaProTyr-ThrProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
Db      793 ---CCTGTGTACAGGAGGGTGCATGGAGAAAGTGCAGGCGGTGGTGCAGGAGAAC 847

```

## RESULT 14

```

US-08-855-140-2
; Sequence 2, Application US/08855140
; Patent No. 5854022
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASCSQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,140
; FILING DATE: Herewith
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0296 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MYOMNOT01
; CLONE: 779308
; US-08-855-140-2

```

```

Alignment Scores:
Pred. No.: 1,36e-15 Length: 977
Score: 232.00 Matches: 60
Percent Similarity: 41.28% Conservative: 30
Best Local Similarity: 27.52% Mismatches: 92
Query Match: 18.08% Indels: 36
DB: 2 Gaps: 6

```

```

US-09-972-970-4 (1-233) x US-08-855-140-2 (1-977)
QY      15 CysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeu 34
Db      140 TCTGCTTGAAGTACATGATGTCCTCTCAATTTGATATTCTGGCTCTGTGGCTGTGGG 199
QY      35 PheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAla 54
Db      200 CTGCTGGGAGTGGCATCTGGCTCTCCGTCGCCAAGCACTTTGCCACCTTCTCC--- 256
QY      55 LeuThrAspLeuGlyCysLeuAspProValTrpLeuPheValValGlyGlyValMet 74
Db      257 ---CCCAGCTTCCCTTCTGTTGTCGAGCCAACTTGTTCATCGCCATAGGCACCATGTC 313
QY      75 SerValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLys 94
Db      314 ATGGTGACGGGCTTCTCGGCTGCTGGGGGCCATCAAGGAAACAAGTGCCTCTCTCCTC 373
QY      95 PhePheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAla 114
Db      374 AGCTTTTTCATGCTCTCTGTTGTCACCTCTCTAGCAGAGCTGATCTTACCCATCTCTTC 433
QY      115 PheValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnValLys 134
Db      434 TTTGTCTATGCAAGGTGAACGAAACGCGCAAGAGGACCTGGAAGGAGCCCTGCTG 493
QY      135 AlaTyr-----ArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyr 152
Db      494 CTGTACCAACACCGAGAACCAACGTGGGGGTGAAGAACGCTTGAACATCATCCAGGCTGAG 553
QY      153 TrpSerCysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCys 172
Db      554 ATGGCATGCTGTGGTGTCACGTACTACACACTGG-----589
QY      173 ThrAspLeuAsnProSerArgGluArgCysGlyValProPheSerCysCysValArgAsp 192
Db      590 -----TACCCAGTGTCTGGGGAGAACACGGTTCCCGACCGCTGTCTGCATG----- 634
QY      193 ProAlaMetSerSerThrProSerValAlaMetMetSerGlySerAsnTrpSerTrpSer 212
Db      635 -----GAGAACTCCAGGCTGC 652
QY      213 SerArgAlaProTyrThrProLysAlaValTrpAlaSerLeuArgSerGlyCys 230

```

Db 653 GGGCGGCAAGCAGCAGCCT-----TTGTGG-----AGAACGGGCTGC 691

## RESULT 15

US-09-016-434-938  
; Sequence 938, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Selhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH

## CLASSIFICATION:

PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0002 US  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 938:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 977 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: MYOMNOT01  
CLONE: 779308

## US-09-016-434-938

Alignment Scores:  
Pred. No.: 1,36e-15 Length: 977  
Score: 232.00 Matches: 60  
Percent Similarity: 41.28% Conservative: 30  
Best Local Similarity: 27.52% Mismatches: 92  
Query Match: 18.08% Indels: 36  
DB: 4 Gaps: 6

## US-09-972-970-4 (1-233) x US-09-016-434-938 (1-977)

Qy 15 CysCysGlyLysTyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeu 34  
Db 140 TGCTGCTTCAAGTACATGATGTTCTCTTCAATTGATATTCTGGCTCTGTGGCTGGG 199  
Qy 35 PheLeuAlaIleGlyLeuTrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAla 54  
Db 200 CTGCTGGGAGTGGCATCTGGCTCTCCGTGTCAGGCAACTTGGCCACCTTCTCC--- 256  
Qy 55 LeuThrAspLeuGlyGlyLeuAspProValTrpLeuPheValValGlyGlyValMet 74  
Db 257 ---CCAGCTTCCCTTCTGCTGACGCCAACCTTGGTTCATCGCCATAGGCACCATTTGTC 313

Qy 75 SerValLeuGlyPheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLys 94  
Db 314 ATGGTGACGGGCTTCTCGGCTGCTGGGGGCGCATCAAGGAAACAAAGTGCCTCTCTTC 373  
Qy 95 PhePheSerValPheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAla 114  
Db 374 AGCTTTTTCATGCTGCTGTTGGTCACCTCTAGCAGAGCTGATCTTACCATCTCTTC 433  
Qy 115 PheValPheLysAspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnValLys 134  
Db 434 TTTGCTACATGCACAAGGTGAACGAGAACGCCAAGAGGACCTGAAGGAAGGCTGCTG 493  
Qy 135 AlaTyr-----ArgAspAspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyr 152  
Db 494 CTGTACCAACCGAGAACCAACGTGGGGCTGAAGAACGCTGGAAACATCATCCAGGCTGAG 553  
Qy 153 TrpSerCysCysGlyAlaArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCys 172  
Db 554 ATGCGATGCTGTGGTCTCACTACACAGACTGG----- 589  
Qy 173 ThrAspLeuAsnProSerArgGluArgCysGlyValProPheSerCysCysValArgAsp 192  
Db 590 -----TACCCAGTGTGGGGGAGAACACACGGTTCCCGACCGCTGTGTCATG----- 634  
Qy 193 ProAlaMetSerSerThrProSerValAlaMetMetSerGlySerAsnTrpSerTrpSer 212  
Db 635 -----GAGAACTCCCGAGGCTGC 652  
Qy 213 SerArgAlaProTyrThrProLysAlaValTrpAlaSerLeuArgSerGlyCys 230  
Db 653 GGGCGCAACGGCACCACGCT-----TTGTGG-----AGAACGGGCTGC 691

Search completed: November 21, 2003, 17:14:15

Job time : 70 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2003, 15:29:18 ; Search time 2152 Seconds  
(without alignments)  
2631.479 Million cell updates/sec

Title: US-09-972-970-4  
Perfect score: 1283  
Sequence: 1 MFGKHQHFQPEVGGCGKYF.....RAPYTPKAYWASLRSGCRTT 233

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q/cpn2.1/USPTO.spool/US09972970/runat\_21112003.125335.27239/app.query.fasta\_1.391  
-DB=EST -QFMT=fastop -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09972970 @CN 1 1 2810 @runat\_21112003.125335.27239 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estro:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_nam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1211.5	94.4	1201	9	AL528502	AL528502 AL528502
2	1169.5	91.2	1494	11	AK008761	AK008761 Mus muscu
3	1163.5	90.7	1194	9	AL529630	AL529630 Mus muscu
4	1158.5	90.3	1005	14	BY708665	BY708665 BY708665
5	1151.5	89.8	849	10	BE615772	BE615772 601279927
6	1151	89.7	938	13	B0856801	B0856801 AGENCOURT
7	1090.5	85.0	901	14	CA454987	CA454987 AGENCOURT
8	1087.5	84.8	781	12	BG770931	BG770931 602719148
9	1083	84.4	1032	12	BM563474	BM563474 AGENCOURT
10	1079.5	84.1	957	9	AL543914	AL543914 AL543914
11	1078.5	84.1	1201	9	AL556793	AL556793 AL556793
12	1071.5	83.5	1028	10	BG477727	BG477727 602521422
13	1066	83.1	718	12	B1914325	B1914325 603182305
14	1024.5	79.9	714	10	BG017161	BG017161 de61f11.Y
15	1023	79.7	902	10	BG478644	BG478644 602525532
16	1020	79.5	970	12	B1909709	B1909709 603070671
17	1018.5	79.4	832	10	BF981395	BF981395 602309182
18	1017.5	79.3	782	12	B1158921	B1158921 602921828
19	1013.5	79.0	1020	12	B1557863	B1557863 603236858
20	1007	78.5	888	13	BX447619	BX447619 BX447619
21	967	75.4	654	12	B1829529	B1829529 603080360
22	966.5	75.3	588	10	B8615323	B8615323 601280719
23	963	75.1	594	10	BF043938	BF043938 BF250022B
24	954	74.4	883	14	CA986012	CA986012 AGENCOURT
25	944.5	73.6	3141	11	AK013350	AK013350 Mus muscu
26	935.5	72.9	753	14	CA315923	CA315923 UI-M-FW0
27	930.5	72.5	899	13	BU911975	BU911975 AGENCOURT
28	905	70.5	538	10	B8683087	B8683087 181563 MA
29	905	70.5	842	13	BQ736112	BQ736112 AGENCOURT
30	895	69.8	758	13	B0320674	B0320674 603852801
31	892.5	69.6	1020	14	BY712345	BY712345 BY712345
32	891	69.4	912	12	B1915506	B1915506 603177449
33	875.5	68.2	933	12	BM475708	BM475708 AGENCOURT
34	874.5	68.2	657	12	B1850548	B1850548 imageqc_1
35	864.5	67.4	960	10	BF119347	BF119347 601757362
36	859.5	67.0	769	10	BF122628	BF122628 601760347
37	853.5	66.5	946	13	BQ720381	BQ720381 AGENCOURT
38	841	65.5	629	13	BU696861	BU696861 LL2in1295
39	839	65.4	685	14	BY734818	BY734818 BY734818
40	836.5	65.2	502	4	BX517485	BX517485 R2PD Mus
41	836	65.2	2108	11	AK033543	AK033543 Mus muscu
42	833.5	65.0	1116	11	AK015705	AK015705 Mus muscu
43	831.5	64.8	1032	12	BM909967	BM909967 AGENCOURT
44	830.5	64.7	797	14	CA321457	CA321457 UI-M-FW0
45	829.5	64.7	1108	13	BQ072716	BQ072716 AGENCOURT

ALIGNMENTS

RESULT 1  
AL528502  
LOCUS  
DEFINITION AL528502 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
ACCESSION AL528502  
VERSION AL528502  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1201)

AL528502 1201 bp mRNA linear EST 23-MAY-2003  
AL528502 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
CDNA clone CSODC029YJ10 5-PRIME, mRNA sequence.

AL528502.2 GI:31066352  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)



# AUTHORS TITLE JOURNAL COMMENT

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
Unpublished  
On Feb 13, 2001 this sequence version replaced gi:12791995.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 3528.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DC029DE05QP1&cluster=3528.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DC029DE05QP1.  
Location/Qualifiers

## FEATURES

source

1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC029YJ10"  
/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 236 a 331 c 340 g 262 t 32 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 3 48e-107 Length: 1201  
Score: 1211 50 Matches: 226  
Percent Similarity: 96.17% Conservatives: 0  
Best Local Similarity: 96.17% Mismatches: 7  
Query Match: 94.43% Indels: 2  
DB: 9 Gaps: 1  
US-09-972-970-4 (1-233) x AL528502 (1-1201)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
DB 172 ATGCCCGGAAGCACCAGCATTTCCAGGAACCTGAGTGGCTGGCTCGGGAATACTTC 231  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
DB 232 CTGTTGGCTTCAACATTGCTTCTGGGTGGTGGAGCCCTGTCTCTGCTATCGGCCTC 291  
QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
DB 292 TGGGCTGGGTGAGAAGGGCTTCTTCGAACATCTCAGCTGACATCTGGAGGC 351  
QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
DB 352 CTTGACCCCGTGGCTGTTGTGGTAGTGGAGGCTCATGTGGTGGCTTGGCT 411  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
DB 412 GAGTGCATTGGGGCCCTCGGAGAACACTTCTGCTCAAGTTTTCTCCGTGTCTTC 471  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
DB 472 GGTCTCATCTTCTCTGAGCTGGCAACAGGGATCTGGCCCTTGTCTTCAAGGACTGG 531  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspLeu 140  
DB 532 ATTCGAGACCACTCAACCTTTCATCAACAACAGCTCAAGGCCTCCGGAGCACATT 591  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160  
DB 592 GACCTCCAGAACTCATTTGCTTGGTTCAGGAATACTGGTCTTCTGGAGSCCCGAGGC 651

QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
DB 652 CCCAATGACTGGAACTCAATATCTACTTCAACTGCACCTGACTGAAACCCACCCGGAG 711  
QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199  
DB 712 CGCTGGGGGTGCCCTTCTCTCTGCTGGCTCAGGACCCCTGGGAGGATGTCTTCAACACC 771  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTrpThrPr 219  
DB 772 CAGTGTGGCTACGACGTCGCGCTCAAACTGGAGCTGGAGCAGCAGGCGCTTATCCACACC 831  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
DB 832 AAAGGCTGGTGGCCAGTTTGAGAAGTGGCTGCAGGACAAACC 874  
AK008761 1494 bp mRNA linear HTC 05-DEC-2002  
Mus musculus adult male stomach cDNA, RIKEN full-length enriched  
library, clone:2210021G21 product:F-BOX PROTEIN FBX23 (FRAGMENT)  
homolog [Homo sapiens], full insert sequence.  
AK008761  
HTC; CAP trapper  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehnenbush, J., Schriml, L.M., Stauber, P., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., But, C.,  
Clatcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hoemann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,



```

VERSION      AL529630.2  GI:31067473
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS      1 (bases 1 to 1194)
              Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      On Feb 13, 2001 this sequence version replaced gi:12793123.
              Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 Evry cedex - France
              Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
              was normalized. Library was constructed by Life Technologies, a
              division of Invitrogen. This sequence belongs to sequence cluster
              3528.f For more information about this cluster, see
              http://www.genoscope.cns.fr/
              cgi-bin/cluster.cgi?seq=CS0DD005DA100P1&cluster=3528.f. Contact :
              Feng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Faraday Avenue Genoscope sequence ID : CS0DD005DA100P1.
              Location/Qualifiers
                1..1194
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="CS0DD005YB20"
                  /tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
                  /clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
                  /note="1st strand cDNA was primed with a NotI-oligo(dT)
                  primer. Five prime end enriched, double-strand cDNA was
                  digested with NotI and cloned into the NotI and EcoR V
                  sites of the pCMVSPORT 6 vector. Library was normalized."
              BASE COUNT  237 a 331 c 336 g 253 t
ORIGIN
Alignment Scores:
Pred. No.:      1 59e-102      Length:      1194
Score:          1163.50      Matches:      223
Percent Similarity: 95.3%      Conservative: 1
Best Local Similarity: 94.8%      Mismatches: 9
Query Match:    90.6%      Indels:      3
DB:              9          Gaps:      1

US-09-972-970-4 (1-233) x AL529630 (1-1194)

QY 1 MetProGlyLysHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20
Db 162 ATGCCCGGCAAGCAGCAGCATTTCCAGGAACCTGAGGTCGGCTGCGGGAATATCTC 221

QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db 222 CTGTTTGGCTTCAACATTGCTCTTCGGGTGGTGGGAGCCCTGTTCTCGCTATCGGCCTC 281

QY 41 TrpAlaTrpGlyGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db 282 TGGGCTGGGTGAGAGGGCTTCTTCGAACATCTCAGCCTGACAGAMCTGGGAGGC 341

QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80
Db 342 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGCGTCATGCTCGGCTTTCGT 401

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
Db 402 GCTGCATTTGGGGCCCTCCGGGAGAACACCTTCCTCGCTCAAGTTTTCCTCGTGTTCCTC 461

QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheIleAspTrp 120
Db 462 GGTCTCATCTTCTCTCGAGCTGGCAAGGGATCTTGGCTTTCCTTCAAGACTGG 521

QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspIle 140

```

```

Db 522 ATTGGAGACAGCTCAACTCTTTCATCAACACACAGCTCAGGCCCTACGGGACGACATT 581
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160
Db 582 AACCTCCAGAACCTCATTCAGCTTTGCTCAGGAATATCGTCTTGTGCGGAGC-CCAGGC 640
QY 161 ProIleAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
Db 641 CCCAATGACTGGAACTCAATATCTACTTCAACTGCATCTGACTTGAACCCCGGAG 700
QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
Db 701 CGCTGGGGGTGCCCTTCTCTGCTGCTCAGGAGCCCTGCGGAGGATGCTCTCAACACC 760
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
Db 761 CAGTGTGCTACGACGCTCCGGCTCAAACTGGAGCGGAGCAGCGGCTTATCCACACC 820
QY 219 olvAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
Db 821 AAAGCTCGTGGGCCAGTTTGAAGAAGTGGCTGCAGGACAACC 863

RESULT 4
LOCUS       BY708665
DEFINITION Mus musculus cDNA clone 2210021G21 5', mRNA sequence.
ACCESSION  BY708665
VERSION     BY708665.1  GI:27119857
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 1005)
AUTHORS    Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
              Nikaido, I., Osato, N., Saito, R., Suzaki, H., Yamanaka, I., Kiyosawa, H.,
              Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schombach, C.,
              Gojbori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,
              Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
              Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chochia, C., Corbani,
              L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,
              A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
              Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
              Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,
              King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
              P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
              H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,
              Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ring,
              Ranachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,
              B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou,
              M., Shinada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,
              R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
              Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa,
              M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
              Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
              M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
              Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
              Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
              K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
              E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
              Analysis of the mouse transcriptome based on functional annotation
              of 60,770 full-length cDNAs
              Nature 420, 563-573 (2002)
              12346683
              22354683
              12466851
              Contact: Yoshihide Hayashizaki
              Laboratory for Genome Exploration Research Group, RIKEN Genomic
              Sciences Center (GSC), Yokohama Institute
              The Institute of Physical and Chemical Research (RIKEN)
              1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              Tel: 81-45-503-9222

```

Fax: 81-45-503-9216  
Email: genome.res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda  
S., Hashizume, W., Hayaehida, K., Hirozane, T., Hori, F., Imotani, K.,  
Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno  
H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,  
Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H.,  
Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,  
Waki, K., Watahiki, A., Muramatsu, M. and Hayaehizaki, Y. Direct  
Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

FEATURES  
source

Location/Qualifiers  
1..1005  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="2210021G21"  
/sex="male"  
/tissue\_type="stomach"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/clone\_lib="RIKEN full-length enriched, adult male  
stomach"  
/note="Site 1: XhoI; Site 2: SstI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGATTCGACTTAAATTAATTAATCCCGCCCCCCC 3']. cDNA  
was cleaved with XhoI and SstI."

BASE COUNT 175 a 281 c 310 g 237 t 2 others

ORIGIN

Alignment Scores:  
Pred. No.: 3 87e-102 Length: 1005  
Score: 1158.50 Matches: 216  
Percent Similarity: 93.59% Conservative: 3  
Best Local Similarity: 92.31% Mismatches: 14  
Query Match: 90.30% Indels: 1  
DB: 14 Gaps: 1

US-09-972-970-4 (1-233) x BY708665 (1-1005)

Qy 1 MetProGlyHisGlnHisPheGlnGluProGluValGlyCysGlyLysTrpPhe 20  
Db 136 ATGCCCGCAAGCAGCAGCAATTCAGGACCTCGAGTGGCTGCTGGGGAATCTTC 195  
Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaGlyLeu 40

Db 196 CTGTTGGCTTCAACATCTGCTTCTGGGTGCTGGAGCCCTGTCTCTGGCCATTGGCCTC 255  
Qy 41 TTPAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 256 TGGGGCTGGGGTGAGAGGGCGTCTCTCCAAATCTCGCGCTGACAGATCTTGGGCGGT 315  
Qy 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 316 CTTGACCCGCTGGCTGCTTGTGGTGGGGGGAGTCTATGTCAAGTGTGGCTTCGCC 375  
Qy 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 376 CGCTGATTGGGGCCCTCCGGGAAACACCTTCTCTCAAGTCTTCTCTGTGTTCTCCTC 435  
Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 436 GGCCTCATCTTCTTCTCGAGCTGGCGGGGATCTCTGGCCCTCTCTTCAAGGATTGG 495  
Qy 121 IleArgAspGlnLeuAsnPhePheIleLeuAsnValLeuAlaTrpArgAspAlle 140  
Db 496 ATCCGAGACCAACTTAATCTTCTTCAACAACAATGTCAAGCCCTACCGGGATGATCTC 555  
Qy 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysGlyAlaArgGly 160  
Db 556 GACCTTCAGAACCTTATCGACTTTCCTCAGGAATACTGCTTGTGTGTCGCCGAGGG 615  
Qy 161 ProAsnAspTrpAsnLeuAsnIleTrpPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 616 CCCAATGATCGAACCTCAATATCTTCACTCACTGACTGACTTGAACCCCAAGCCGTGAG 675  
Qy 181 ArgCysGlyValProPheSerCysValArgAspPro--AlaMetSerSerThrPro 199  
Db 676 CGCTGTGGGTGGCTTCTTCTGCTGTGTAGGACACGCGGGAAGATGCTCAATACCC 735  
Qy 200 SerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTrpThrPro 219  
Db 736 AGTGTGGCTATGACATCCGACTCAAACTGGAGCTCGAGCAGGAGGGCTCATCTACACCA 795  
Qy 220 LysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
Db 796 AAGCTGTGTGGCCAGTTTGAGAAGTGGCTACAAAGAACACC 837  
RESULT 5  
BE615772  
LOCUS BE615772.1 GI:9897371  
DEFINITION BE615772.1 Homo sapiens mRNA linear EST 24-AUG-2000  
mRNA sequence.  
ACCESSION BE615772  
VERSION BE615772.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 849)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM296 row: e column: 08  
High quality sequence stop: 699.  
Location/Qualifiers  
1..849  
/organism="Homo sapiens"

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3621871"
/tissue_type="adenocarcinoma"
/lab_host="NIH MGC 39"
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dr priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT 165 a 241 c 243 g 199 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1.49e-101 Length: 849
Score: 1151.50 Matches: 218
Percent Similarity: 93.59% Conservative: 1
Best Local Similarity: 93.16% Mismatches: 14
Query Match: 89.75% Indels: 2
DB: 10 Gaps: 1

US-09-972-970-4 (1-233) x BE615772 (1-849)
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20
Db |||||
QY 33 ATGCCCGGCAAGCAGCAGCATTTCCAGGAACCTTGAGTGGCTGGCTGGCGGAAAC-ATTC 91
Db |||||
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db |||||
QY 92 CTGTTGGCTTCAACATTTCTTGGGTGCTGGAGCCCTGTCTCGCTATCGGCCTC 151
Db |||||
QY 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGly 60
Db |||||
QY 152 TGGGCTGGGGTGAGAAGGGCGTTCTCTCGAATCTCAGCGCTGACAGATCTGGGAGGC 211
Db |||||
QY 61 LeuAspProValTrpLeuPheValValGlyValGlyValMetSerValLeuGlyPheAla 80
Db |||||
QY 212 CTTGACCCCGTGGCTGTTTGGTGTAGTTGAGGGCGTCATGCTGGTGTGGGCTTGGCT 271
Db |||||
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
Db |||||
QY 272 GCGTCGATTGGGGCCCTCGGAGAACACCTTCTCGTCAAGTCTTCTCGCTGTTCTC 331
Db |||||
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
Db |||||
QY 332 GGTCTCATCTCTTCTGGAGCTGGCAACAGGGATCTGGCCCTTTGTCTTCAAGGACTGG 391
Db |||||
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle 140
Db |||||
QY 392 ATTCGAGACCACTCAACTCTTCATCAACACACAGCTCAAGCCCTACCGGACGACATT 451
Db |||||
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160
Db |||||
QY 452 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACTGCTGTCTGGGAGCCCGAGGC 511
Db |||||
QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
Db |||||
QY 512 CCCAATGACTGGAACTCAATATCTACTTCAACTGCACCTGACTTGAACCCCGAGCGGAG 571
Db |||||
QY 181 ArgCysGlyValProPheSerCysValArgAspProAlaMetSer---SerThrPro 199
Db |||||
QY 572 CGCTGGGGGTGCCCTTCTCTGCTGCTCAGGACCTTGGGAGGATGTCCTCAAAACCC 631
Db |||||
QY 200 SerValAlaMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyrTrpPro 219
Db |||||
QY 632 AGTGTGGCTACGACGCTCGGGCTCAAACTGGAGCTGGAGCTGGAGCAGCGGCTTCATCCACAA 691
Db |||||
QY 220 LysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
Db |||||
```

```
Db 692 AAGCGTGGTGGCCAGTCTGAGAGTGGCTGGCAGGAACC 733

RESULT 6
BU856801
LOCUS
DEFINITION
AGENCOURT 10475185 NIH_MGC_107 Homo sapiens cDNA clone
IMAGE:6646576 5', mRNA-sequence.
BU856801
VERSION
BU856801.1 GI:24041791
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 938)
NIH-MGC http://mgc.mci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2885 row: j column: 16
High quality sequence stop: 649.
FEATURES
source
1..938
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6646576"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/notes="Organ: breast; Vector: pOTB7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dr priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library."
BASE COUNT 162 a 290 c 279 g 207 t
ORIGIN

Alignment Scores:
Pred. No.: 1.89e-101 Length: 938
Score: 1151.00 Matches: 218
Percent Similarity: 93.22% Conservative: 2
Best Local Similarity: 92.37% Mismatches: 13
Query Match: 89.71% Indels: 3
DB: 13 Gaps: 2

US-09-972-970-4 (1-233) x BU856801 (1-938)
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20
Db |||||
QY 113 ATGCCCGGCAAGCAGCAGCATTTCCAGGAACCTTGAGTGGCTGGCGGAAACATCTTC 172
Db |||||
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db |||||
QY 173 CTGTTTGGCTTCAACATTTCTTGGGTGCTGGAGCCCTGTTCTCGCTATCGGCCTC 232
Db |||||
QY 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGly 60
Db |||||
QY 233 TGGGCTGGGGTGAGAAGGGCGTTCTCTCGAATCTCAGCGCTGACAGATCTGGGAGGC 292
Db |||||
QY 61 LeuAspProValTrpLeuPheValValGlyValMetSerValLeuGlyPheAla 80
Db |||||
```

Db 293 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGCGTCATGTCGGTGTGGCTTTGCT 352

QY 81 GIVCystileGlyAlaLeuArgGluAenThrPheLeuLeuLysPheSerValPheLeu 100

Db 353 GGCTGCATTTGGGCGCCCTCCGGAGAACACCTTCTGCTCAAGTTTTTCTCCGTGTTCCCTC 412

QY 101 GlyLeuilePhePheLeuGluLeuAlaThrGlyileLeuAlaPheValPheLysAspTrr 120

Db 413 GGTCTCATCTTCTCTGGAGCTGGCAACAGGATCTCTGGCCTTTGTCTTCAAGGACTGG 472

QY 121 IleArgAspGlnLeuAenPhePheIleAsnAenVallyAlaTyArgAspAspIle 140

Db 473 ATTCGAGACAGCTCAACCTTCTCATCAACAACCAACCTCAAGGCTTACCGGACGACATT 532

QY 141 AspLeuGlnAenLeuileAspPheAlaGlnGluTyTrpSerCysGlyAlaArgGly 160

Db 533 GACCTCAGAACCTCATTTGCTTGTCTCAGGAATACTGGTCTTGTCTGGAGCCCGAGGC 592

QY 161 ProAsnAspTrpAsnLeuAenIleTyPheAsnCysThrAspLeuAenProSerArgGlu 180

Db 593 CCCAATGACTGGAACTCAATATCTACTTCAACTGCACTGACTTGAACCCCGCGGAG 652

QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199

Db 653 CGCTGGCGGTGGCCCTTCTCTGCTGCTCAGGACCCCTCGGAGGATGTCCCTCACACC 712

QY 199 oSerValAlaMetMetSerGlySerAenTrpSerTrpSerSerArgAlaProTyThrPr 219

Db 713 CAGTGTGGCTACAGCTCCGGCTCAAACTGGAGCTGGAGCAGCAAGGCTTCATCCACACC 772

QY 219 oLysAlaValTrpAlaSerLeu---ArgSerGlyCysArgThrThr 233

Db 773 AAAGGCTGGCGGGCCCCAAATTGAAAGAGTGGCTGCAAGGACCCACC 818

## RESULT 7

CA454987 901 bp mRNA linear EST 12-NOV-2002  
 AGENCOURT 10735735 MAPcL Homo sapiens cDNA clone IMAGE:6722499 5',  
 mRNA sequence.

CA454987.1 GI:24905277

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Homosapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@rs@mail.nih.gov  
 Tissue Procurement: Kristi A. Eglund, Ira Pastan  
 cDNA Library Preparation: Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LILNL at:  
 http://image.lnl.gov  
 Place: LILNL4285 row: f column: 03  
 High quality sequence stop: 683.  
 Location/Qualifiers  
 1. .901  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6722499"  
 /cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HME1  
 , LNCap"  
 /lab\_host="EMDH108"  
 /clone\_lib="MASCN"  
 /note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;  
 Subtracted with brain, liver, lung, kidney and muscle.

## FEATURES

## source

BASE COUNT 157 a 275 c 269 g 200 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,35e-95 Length: 901  
 Score: 1090.50 Matches: 206  
 Percent Similarity: 93.69% Conservative: 2  
 Best Local Similarity: 92.79% Mismatches: 11  
 Query Match: 85.00% Indels: 3  
 DB: 14 Gaps: 1  
 US-09-972-970-4 (1-233) x CA454987 (1-901)  
 QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyPhe 20  
 Db 133 ATCCCGCGCAAGCACCAGCATTTCAGGAACTTCAGGTCGGCTGCGGGAATACTTC 192  
 QY 21 LeuPheGlyPheAenIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 193 CTGTTTGGCTTCAACATTGTCTTCTGGTGTCTGGAGCCCTGTCTTGGCTATTCGGCTC 252  
 QY 41 TrpAlaTrpGlyGlnLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 253 TGGCGCTGGGTGAGAAGGCGTCTCTCGAATCTCAGCGCTGCAGATCTGGGAGGC 312  
 QY 61 LeuAspProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80  
 Db 313 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGGAGGCTCATGTCTGGTGTGGCTTTGCT 372  
 QY 81 GlyCysileGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePheSerValPheLeu 100  
 Db 373 GGCTGCATTTGGGCGCCCTCCGGAGAACACCTTCTGCTCAAGTTTTTCTCCGTGTTCCCTC 432  
 QY 101 GlyLeuilePhePheLeuGluLeuAlaThrGlyileLeuAlaPheValPheLysAspTrr 120  
 Db 433 GGTCTCATCTTCTCTGGAGCTGGCAACAGGATCTCTGGCCTTTGTCTTCAAGGACTGG 492  
 QY 121 IleArgAspGlnLeuAenPhePheIleAsnAenVallyAlaTyArgAspAspIle 140  
 Db 493 ATTCGAGACAGCTCAACCTTCTCATCAACAACCAACCAAGGCTTACCGGACGACATT 552  
 QY 141 AspLeuGlnAenLeuileAspPheAlaGlnGluTyTrpSerCysGlyAlaArgGly 160  
 Db 553 GACCTCAGAACCTCATTTGCTCAGGAATACTGGTCTTGTCTGGAGCCCGAGGC 612  
 QY 161 ProAsnAspTrpAsnLeuAenIleTyPheAsnCysThrAspLeuAenProSerArgGlu 180  
 Db 613 CCCAATGACTGGAACTCAATATCTACTTCAACTGCACTGACTTGAACCCCGCGGAG 672  
 QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199  
 Db 673 CGCTGGCGGTGGCCCTTCTCTGCTCAGGACCCCTCGGAGGATGTCTCTCAACAAC 732  
 QY 199 oSerValAlaMetMetSer-GlySerAenTrpSerTrpSerSerArgAlaProTyThrPr 219  
 Db 733 CAGTGTGGCTACAGCTCCGAGCTTCACTGGAGCTGGGAAACAACAGGCTTTCATCCCC 792  
 QY 219 IO 219  
 Db 793 CA 794  
 RESULT 8  
 BG770931  
 LOCUS  
 DEFINITION  
 mRNA sequence.  
 ACCESSION  
 BG770931

BG770931 781 bp mRNA linear EST 15-MAY-2001  
 602719148F1 NTH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4839779 5',  
 mRNA sequence.  
 ACCESSION  
 BG770931

Directly cloned. Priming method: oligo-dT. Average  
 insert size: 1800 bp. Library amplification: 26,000 fold.  
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,  
 Bungkok Lee & Ira Pastan: Discovery of new breast  
 cancer genes encoding membrane and secreted proteins.  
 Manuscript submitted."



```

VERSION BG770931.1 GI:14081584
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 781)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
COMMENT NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI669 row: o column: 12
High quality sequence stop: 697.
Location/Qualifiers
1..781
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4839779"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC 60"
/notes="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site 1: SfII (ggccgctggcc); Site 2: SfII (ggccattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-DT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT 127 a 239 c 239 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 2,19e-95 Length: 781
Score: 1087.50 Matches: 203
Percent Similarity: 94.88% Conservative: 1
Best Local Similarity: 94.42% Mismatches: 10
Query Match: 84.76% Indels: 2
DB: 12 Gaps: 1

US-09-972-970-4 (1-233) x BG770931 (1-781)
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20
Db 129 ATGCCCGGACAGCAGCATTCACAGAACCTGAGTGGTGGTGGCGGAAATCTTC 188
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db 189 CTGTTGGCTTCAACATTCCTCTGGGTGGTGGGAGCCCTGCTCTGGCTATCGCCCTC 248
QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db 249 TGGGCTGGGGTGAGAGGGCGTCTCTCGAACATCTCAGCCCTCAGCATCTGGGAGGC 308
QY 61 LeuAspProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80
Db 309 CTTGACCCCGTGTGGCTGTTTGTGGTAGTTGAGGCGCATCTCGTGGTGGGCTTGGCT 368
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100

```

```

Db 369 GGCTGCATTGGGCCCCCTCGGGAGAACACCTTCTCTCAAGTTTTTCTCCGTGTTCTCCTC 428
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
Db 429 GGTCTCATCTTCTTCTGAGCTGGCAACAGGATCTGGCCCTTGTCTTCAAGGACTGG 488
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValIysAlaTyrArgAspAspIle 140
Db 489 ATTCTGAGACAGCTCAACCTCTTCATCAACAACAGCTCAAGGCCCTACCGGGACGACATT 548
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyValaIaArgGly 160
Db 549 GACCTCCAGAACCTCATTCATCTTGTCTCAGGATACTGGTCTTCTGGGAGCCGAGGC 608
QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
Db 609 CCCAATGACTGGAACTCAATATCTACTTCAACTGCACTGACTTGAACCCAGCCGGGAG 668
QY 181 ArgCysGlyValProPheSerCysValArgAspProAlaMetSerSer---ThrPro 199
Db 669 CGCTGGGGGTGCCCTTCTCTGCTCGTCCAGAACCCCTCGCGAAGGATGTCCTCAAAACC 728
QY 200 SerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArg 214
Db 729 AGTGTGCTACGACGTCCGGTCA-AACTGGAGCTGGAGGCAGAGG 772

RESULT 9
BM563474 1032 bp mRNA linear EST 20-FEB-2002
LOCUS AGENCOURT_6589677 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441687
5' mRNA sequence.
ACCESSION BM563474
VERSION BM563474.1 GI:18810427
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI916 row: f column: 24
High quality sequence stop: 697.
Location/Qualifiers
1..1032
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5441687"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_98"
/notes="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 203 a 309 c 286 g 232 t 2 others
ORIGIN

```

Alignment Scores: 8.62e-95 Length: 1032  
 Pred. No.: 1083.00 Matches: 204  
 Score: 86.81% Conservative: 0  
 Percent Similarity: 86.81% Mismatches: 6  
 Best Local Similarity: 84.41% Indels: 25  
 Query Match: 12 Gaps: 2  
 DB:

US-09-972-970-4 (1-233) x BM563474 (1-1032)

Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20  
 Db 56 ATGCCCGCAAGCAGCAGCACTTCAGGAACCTGAGTGGCTGCTGGCGGAATACCTTC 115

Qy 21 LeuPheGlyPheAenIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 116 CTGTTGGCTTCAACATTTCTTCTGGGTGCTGGAGCCCTTCTTCTGGCTATCGGCCTC 175

Qy 41 TrpAlaTrpGlyGluLysGlyValLeuSerAenIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 176 TGGGCTGGGTGAGAGGGCTTCTCTCGAACATCTCAGCGCTCAGAGATCTGGGAGGC 235

Qy 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
 Db 236 CTTGACCCCGTGGCTGCTTTGTGTAGTGTGGAGCGCTCATGTGGTCTGGGCTTGTCT 295

Qy 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
 Db 296 GGCTGCATTTGGGGCCCTCCGGAGAACACCTTCTGCTCAAGTTT----- 340

Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
 Db 341 -----GACTGG 346

Qy 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle 140  
 Db 347 ATTCGAGACCAAGCTCAACTCTTATCAACACACACGCTCAAGGCGCTACCGGAGCAGATT 406

Qy 141 AspLeuGlnAenLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyValAArgGly 160  
 Db 407 GACTCCAGAACCTCATTTGCTCAGGAATACTGGTTCTTCTGGAGCCCGGAGGC 466

Qy 161 ProAsnAspTrpAenLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
 Db 467 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCAGCTGACTGAAACCCAGCCGGAG 526

Qy 181 ArgCysGlyValProPheSerCysValArgAspProAla-----MetSerSerThrPr 199  
 Db 527 CGCTGCGGGTGGCTTCTCTGCTGCTCAGGACCCCTGCGAGGATGCTCTCAACACC 586

Qy 199 oSerValAlaMetMetSerGlySerAenTrpSerTrpSerSerArgAlaProTyrThrPr 219  
 Db 587 CAGTGTGCTACGAGTCCGGCTCAAACTGGAGCTGGAGAGCGAGGCTTCAACACACC 646

Qy 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
 Db 647 AAAGGCTGCGTGGGCCAGTTTGAGAGTGGCTGCGAGACCAACC 689

RESULT 10  
 AL543914 957 bp mRNA linear EST 31-MAY-2003  
 LOCUS AL543914 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 DEFINITION clone CSODI005YM07 5-PRIME, mRNA sequence.  
 ACCESSION AL543914.2 GI:31265759  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 957)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization

Unpublished  
 On Feb 15, 2001 this sequence version replaced gi:12876393.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 3528.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSODI005AG04QPI&cluster=3528.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSODI005AG04QPI.

FEATURES  
 Location/Qualifiers  
 1..957  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODI005YM07"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="First strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 BASE COUNT 176 a 279 c 288 g 205 t  
 ORIGIN

Alignment Scores: 1.71e-94 Length: 957  
 Pred. No.: 1079.50 Matches: 208  
 Score: 88.19% Conservative: 1  
 Percent Similarity: 87.76% Mismatches: 12  
 Best Local Similarity: 84.14% Indels: 16  
 Query Match: 9 Gaps: 2  
 DB:

US-09-972-970-4 (1-233) x AL543914 (1-957)

Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20  
 Db 172 ATGCCCGCAAGCAGCAGCACTTCAGGAACCTGAGTGGCTGCTGGCGGAATACCTTC 231

Qy 21 LeuPheGlyPheAenIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 232 CTGTTGGCTTCAACATTTCTTCTGGGTGCTGGAGCCCTTCTTCTGGCTATCGGCCTC 291

Qy 41 TrpAlaTrpGlyGluLysGlyValLeuSerAenIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 292 TGGGCTGGGTGAGAGGGCTTCTCTCGAACATCTCAGCGCTCAGAGATCTGGGAGGC 351

Qy 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
 Db 352 CTTGACCCCGTGGCTGCTTTGTGTAGTGTGGAGCGCTCATGTGGTCTGGGCTTGTCT 411

Qy 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
 Db 412 GGCTGCATTTGGGGCCCTCCGGAGAACACCTTCTGCTCAAGTTT----- 456

Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
 Db 457 -----CTGGCAACAGGATCTTGGGCTTGTCTTCAAGACTGG 495

Qy 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle 140  
 Db 496 ATTCGAGACCAAGCTCAACCTTCTCATCAACACACGCTCAAGGCTACCGGAGCAGCATT 555

Qy 141 AspLeuGlnAenLeuIleAspPheAlaGln-----GluTyrTrpSerCysCysGlyAlaArg 160  
 Db 556 GACCTCCAGAACCTCATTTGCTTGTGTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 615

Qy 160 lyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArg 180



```

Db      616 CCCTCAATGACTGGAACCTCAATATCTACTTCAACTGCACCTGACTTGAACCCCGCGG 675
Qy      180 luAryCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThr 198
Db      676 AGCGCTGGGGTGCCTTCTCTCTGCTGCGTCAGGACCCCTCGGAGGATGTCCTCAACA 735
Qy      199 ProSerValAlaMetMetSerGlySerAntTTPSerTTPSerSerArgAlaProTyrThr 218
Db      736 CCCAGTGTGGCTACGACGTCCTGGCTCAAACTGGAGCTGGAGCAGGAGGCTTCATCCACA 795
Qy      219 ProLysAlaValTTPAlaSerLeuArgSerGlyCysArgThrThr 233
Db      796 CCAAGAGCTGCGTGGCCAGTTTGAGAGTGGCTGCGAGCAACC 840

RESULT 11
AL556793
LOCUS      AL556793 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DK005YH21 5-PRIME, mRNA sequence.
ACCESSION AL556793
VERSION    AL556793.2 GI:31278594
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 1201)
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
COMMENT    Full-length cDNA libraries and normalization
            On Feb 15, 2001 this sequence version replaced gi:12899797.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DK005CD11QPI&cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DK005CD11QPI.
            Location/Qualifiers
                1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0DK005YH21"
                /cell_type="HELA CELLS COT 25-NORMALIZED"
                /cell_line="HELA"
                /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
                /note="1st strand cDNA was primed with a NotI-oligo(dT)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and cloned into the Not I and EcoR V
                sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      231 a 322 c 324 g 258 t 66 others
ORIGIN

Alignment Scores:
Pred. No.:      2,876-94      Length:      1201
Score:          1078.50      Matches:      209
Percent Similarity: 92.54%      Conservative: 2
Best Local Similarity: 91.67%      Mismatches: 11
Query Match:      84.06%      Indels: 7
Db:              9          Gaps: 3

US-09-972-970-4 (1-233) x AL556793 (1-1201)

Qy      1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20
Db      268 ATGCCCGCAAGCACCACATTTCCAGGAACCTTGAGGTGCGTCTGCGGGAATACTTTC 327

```

```

Qy      21  LeuPheGlyPheAsnIleValPheTTPValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
Db      328 CTGTTTGGCTTCAACATTTGCTTCTGGGTGCTGGGAGCCCTGTTCTTGGCTATCGGCCTC 387
Qy      41  TrpAlaTTPGlyGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
Db      388 TGGGCTCGGGGTGAGAAGGGCTTCTCTCGAACATCTCAGCGCTCAGAGATCTGGGAGGC 447
Qy      61  LeuAspProValTTPLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80
Db      448 CTTGACCCCGTGTGGCTGTTTGTGTAGTTGGAGCGCTCATGTCGTTGCTTGGGCTTTGCT 507
Qy      81  GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
Db      508 GCTGCAATGGGGCCCTCCGGAGAACACCTTCTCTGCTCAAGTTTCTCCGTGTCTCTC 567
Qy      101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTTP 120
Db      568 GGTCTCATCTTCTTCTTCTGGAGCTGGCAACAGGATCTTGGCTTGTCTTCAAGGACTCG 627
Qy      121 IleArgAspGlnLeuAsnPhePheIleAsnValLysAlaTyrArgAspAspIle 140
Db      628 ATTGAGACCACTCAACCTCTTTCATCAACAACAACAGTCAAGGCTTACCGGAGCAGCAT 687
Qy      141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTTPSerCysCysGlyAlaArgGly 160
Db      688 GACTCCAGAACCTCATTTGCTCAGGAATACTGGTCTTGTGCGGAGSC-CGAGGC 746
Qy      161 ProAsnAspTTPAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
Db      747 CCCAATGACTGGAACTCAATATCTACTTCAACTGCATCTGACTTGAACCCAGCCGGAG 806
Qy      181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
Db      807 CGCTCGGGGTGCCCTTCTCTGCTGCTCAGGAGCCCTGCGGAGGATKTCCTCAACACC 866
Qy      199 oSerValAlaMetMetSerGlySerAsnTTP-----SerTTPSerSerArgAlaPr 216
Db      867 CAGTGTGGCTACGACGTCGGCTCAACTGGTGAGAGGGAGCTKGASAGCAGGSGSYTAT 926
Qy      216 oTyrThrProLysAlaValTTP 223
Db      927 YCACACA---AAGCTGGGTGG 945

RESULT 12
BG477727
LOCUS      BG477727 1028 bp mRNA linear EST 21-MAR-2001
DEFINITION 602521422F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:463997 5',
            mRNA sequence.
ACCESSION  BG477727
VERSION     BG477727.1 GI:13410006
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 1028)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC/BCRP/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM1404 row: k column: 06
            High quality sequence stop: 734.
FEATURES
            Location/Qualifiers
                1..1028

```

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4639997"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 20"
/notes="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 194 a 236 c 310 g 228 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 1,12e-93 Length: 1028
Score: 1071.50 Matches: 214
Percent Similarity: 91.10% Conservative: 1
Best Local Similarity: 90.68% Mismatches: 17
Query Match: 83.52% Indels: 6
DB: 10 Gaps: 1
```

US-09-972-970-4 (1-233) x BG477727 (1-1028)

```
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20
DB 134 ATCCCGCGGCAAGCACCACGATTTCCAGGAACCTGAGTGGCTGCTCGCGGAATACTTC 193
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyValAlaLeuPheLeuAlaIleGlyLeu 40
DB 194 CTGTTTGGCTTCAACATTTGTTCTGGTCTGGAGCCCTGTTCTCTGGCTATCGGCCTC 253
QY 41 TrpAlaTrpGlyGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
DB 254 TGGGCTGGGGTGAAGGCGGTTCTCTCGAACAATCTCAGCGCTGACAGATCTGGGAGGC 313
QY 61 LeuAspProValTrpLeuPheValValValGlyValMetSerValLeuGlyPheAla 80
DB 314 CTTGACCCCGTGGGCTGTTTGTGGTAGTTGGAGGCGTCATGTGGTGTGGCTTTGCT 373
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPheSerValPheLeu 100
DB 374 GGCTGATTTGGGCGCTTCGGGAGAACCTTCTCTCAAGTTTTTCTCCGTGTTCTTC 433
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
DB 434 GGTCCTCATCTTCTCTGGAGCTGGCAACAGGGATCTGGCCTTTGCTTCAAGGACTGG 493
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValIlyAlaIleValArgAspIle 140
DB 494 ATTCGAGACGAGCTCAACCTTTCATCAACAACAGCTCAAGGCTTACCGCGACGAT 553
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysGlyAlaArgGly 160
DB 554 GACCTCCAGAACCTCATTTGCTTGGAGTAATCTGGTCTGTGGAGCCCGAGG- 612
QY 161 ProAsnAspTrpAsnLeuAsnIleValPheAsn-CysThrAspLeuAsnProSerArgG 180
DB 613 CCCAATGACTGGAACTCAATATCTACTTCAACTTGCATCTGACTTGAACCCCGCGGA 672
QY 180 uArgCysGlyValProPheSerCysValArgAsp---ProAlaMetSerSerThrPr 199
DB 673 GCGCTCGGGGGGCGCTTCTCTGCTGCGTAAGGACCTCGCGGAGGATGTCTCAAAACC 732
QY 199 oSerVal-AlaMetMetSerGlySerAsnTrp-SerTrpSerSerArgAlaProTyrThr 218
DB 733 CAGTGTGGGCTACACGCTCCGCTCAACTGGAGCTGGAGCGAGGGGTTTCATCCACA 792
QY 219 ProLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232
```

```
Db 793 CCAAGGT-GCTGGGCCAGTTGAAAAGTGGCTTGCAAGGACA 833
RESULT 13
BI914325
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
```

```
BASE COUNT 115 a 222 c 217 g 164 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 2,4e-93 Length: 718
Score: 1066.00 Matches: 193
Percent Similarity: 99.48% Conservative: 0
Best Local Similarity: 99.48% Mismatches: 1
Query Match: 83.09% Indels: 0
DB: 12 Gaps: 0
```

US-09-972-970-4 (1-233) x BI914325 (1-718)

```
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20
DB 119 ATGCCCGGCAACACGACGATTTCCAGGAACCTGAGTGGCTGCTCGCGGAATACTTC 178
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyValAlaLeuPheLeuAlaIleGlyLeu 40
DB 179 CTGTTTGGCTTCAACATTTGTTCTGGTCTGGAGCCCTGTTCTCTGGCTATCGGCCTC 238
QY 41 TrpAlaTrpGlyGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
DB 239 TGGGCTGGGGTGAAGGCGGTTCTCTCGAACAATCTCAGCGCTGACAGATCTGGGAGGC 298
```

```
||||| ||||||| ||||||| ||||||| |||||||
793 CCAAGGT-GCTGGGCCAGTTGAAAAGTGGCTTGCAAGGACA 833
BI914325
603182305F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246304 5',
mRNA sequence.
718 bp mRNA linear EST 17-OCT-2001
BI914325
BI914325.1 GI:16198834
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 718)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
DNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11621 row: j column: 01
High quality sequence stop: 712.
Location/Qualifiers
1..718
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246304"
/lab_host="DH10B"
/clone_lib="NIH MGC_121"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 017. Note:
this is a NIH MGC Library."
```

QY 61 LeuAspProValTrrPheValValValGlyValMetSerValLeuGlyPheAla 80  
 Db 299 CTTGACCCCGTGGCTTTGGTAGTTGGAGCGTCATGTCGGGTGGGCTTGGCT 358  
 QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
 Db 359 GGCTGCAATGGGCGCTCCGGAGAACACCTTCTGCTCAAGTTTTTCTCCGTGTTCTC 418  
 QY 101 GlyLeuIlePhePheLeuGluAlaThrGlyIleLeuAlaPheValPheLysAspTrr 120  
 Db 419 GGCTCATCTCTCTCCGGAGCTGGCAACAGGATCCTGGCCCTTGTCTTCAAGGACTGG 478  
 QY 121 IleArgAspGlnLeuAsnThrPheIleAsnAsnValLysAlaValArgAspAspIle 140  
 Db 479 ATTCGAGACAGCTCAACCTTTCATCAACAACAGCTCAAGGCCTACCGGACGACATT 538  
 QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrrPheSerCysGlyAlaArgGly 160  
 Db 539 GACCTCAGAACCTCATGTGCTTGTCTCAGGAATAGTGTCTTGTCTGGAGCCCGAGGC 598  
 QY 161 ProAsnAspTrrPheAsnLeuIleTrrPheAsnThrAspLeuAsnProSerArgGlu 180  
 Db 599 CCAATGACTGGAACTCAATATCTACTTCACTGCACTGCACTGAAACCCCGCGGAG 658  
 QY 181 ArgCysGlyValProPheSerCysValArgAspProAla 194  
 Db 659 CGCTGGGGGTGCCCTTCTCTGCTGGTCAGGAGCCCTGCG 700

## RESULT 14

BG017161

LOCUS

DEFINITION

des611.1 y1 Kirschner embryo St10 14 Xenopus laevis cDNA clone  
 IMAGE:3516428 5' similar to TR:060628 060628 TETRASPANIN TSPAN-5.

[1] ; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Washu Xenopus EST project, 1999  
 Unpublished  
 Other ESTs: des611.x1  
 Contact: Sandy Clifton, Ph.D.  
 Washu Xenopus EST project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Library constructed by M. Kirschner (Harvard Medical School). DNA  
 Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: Xenopus clones from this library are available  
 through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 442.  
 Location/Qualifiers  
 1..714  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:3516428"  
 /tissue\_type="pooled embryos (stage 10-14)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="Kirschner embryo St10 14"  
 /note="Vector: pCS2+; Site\_1: NotI; Site\_2: SalI;

FEATURES

Source

RESULT 15

BG478644

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

BG478644  
 60525532F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:4643959 5',  
 mRNA sequence.  
 BG478644  
 BG478644.1 GI:13410923  
 EST.  
 Homo sapiens (human)

Size-selected for average insert size 1.2 kb. Library was  
 constructed and donated by M. Kirschner (Harvard Medical  
 School)."

BASE COUNT 170 a 154 c 177 g 212 t 1 others  
 ORIGIN

Alignment Scores:

Pred. No.: 2,56e-89 Length: 714  
 Score: 1054.50 Matches: 190  
 Percent Similarity: 85.84% Conservative: 10  
 Best Local Similarity: 81.55% Mismatches: 31  
 Query Match: 79.85% Indels: 2  
 DB: 10 Gaps: 1

US-09-972-970-4 (1-233) x BG017161 (1-714)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 Db 10 ATGTACGCAACATCAGCATTTTAAAGGTCCCGAAGTGAGCTGCTGTGTGAAGATTTC 69  
 QY 21 LeuPheGlyPheAsnIleValPheTrrPheValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 70 TTGTTTGGATTCAACATAGTGTTTTGGTTGCTTGGAGCTAGTTTCTAGCTATTGGCTTT 129  
 QY 41 TrrAlaTrrPheGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 130 TGGCATGGCGGAGAAAGGGTGTCTGTCCACATCTACTCCATCAGACCTTGGTGGC 189  
 QY 61 LeuAspProValTrrPheValValValGlyValMetSerValLeuGlyPheAla 80  
 Db 190 TTTGACCCCGTTTGGCTCTTCATAGTCACTGGAGCTGTCTATGTTGTGGGGTTGGC 249  
 QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
 Db 250 GGCTGCATTGGAGCCTTGAGAGAAACACGACACTCTTAAATTTTCTCAGTCTTCCCTG 309  
 QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrr 120  
 Db 310 GGGTGTATCTTCTTGGAGCTGACTGCGAGGATCTTGGCCCTTGTCTTCAAGACTGG 369  
 QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaValArgAspAspIle 140  
 Db 370 ATCAAAGATCAGCTTAATTTTTTATCAACAACACGTCGAGCGTATCGTGATGATTT 429  
 QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrrPheSerCysGlyAlaArgGly 160  
 Db 430 GACCTCCAGAACCTCATTTGCTCAGGAATATTGGTCTGCTGCTGTCGACATGGA 489  
 QY 161 ProAsnAspTrrPheAsnIleTrrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
 Db 490 CCTAACGACTGGAACTCAACATCTACTTAACTGTACAGACTCTTAACCCAGCAGGAA 549  
 QY 181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199  
 Db 550 CGCTGGGTGTGCCATTCTCTGCTGTGTGAAGGACCTTCCGAGAGATGTTCCACACACA 609  
 QY 199 oSerValAlaMetMetSerGlySerAsnTrrPheSerSerArgAlaProTrrThrPr 219  
 Db 610 CAGTGTGTATATGATGTGGCTGAAGCTGGAAGTCTGGAGCAGCATAGCTTCTATACC 669  
 QY 219 oLysAlaValTrrAlaSerLeuArgSerGlyCysArg 231  
 Db 670 AAAGGATGTGTGGCCCGAGTGTGAGCAGTGGCTGCAAG 706

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 902)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-2@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L141414 row: p column: 08  
High quality sequence stop: 813.  
Location/Qualifiers  
1..902  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4643959"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 20"  
/notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5',  
adaptor: GGACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 146 a 268 c 288 g 200 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4, 85e-89 Length: 902  
Score: 1023.00 Matches: 210  
Percent Similarity: 88.61% Conservatives: 0  
Best Local Similarity: 88.61% Mismatches: 19  
Query Match: 79.73% Indels: 9  
DB: 10 Gaps: 1  
US-09-972-970-4 (1-233) x BG478644 (1-902)  
QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
Db 136 ATGCCCGGCAAGCACCAGCATTTCCAGGAACCTGAGTCGGTCTGCTGGGAAATATCTTC 195  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 196 CTGTTTGGCTTCAACATTGCTCTTGGGTGCTGGGAGCCCTGTTCTCGGTATATCGGCCTC 255  
QY 41 TrpAlaTrpGlyGlyLeuSerValLeuSerAsnIleSerAlaLeuThrAsp-LeuGlyG1 60  
Db 256 TGGGCTTGGGTGAGAGGGCGTTCTCTGAACATCTCAGCGCTCAGAGATCTCGGGAGG 315  
QY 60 YLeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAl 80  
Db 316 CCTTGACCCCGTGGCTGTTTGTGGTAGTTGGAGCGCTCATGTCGGTGTGGGCTTTGC 375  
QY 80 aGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLe 100  
Db 376 TGGCTGCATTGGGGCCCTCCGGGAGAACACCTTCTGCTCAAGATTTTCTCGGTGTTCT 435  
QY 100 uGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120  
Db 436 CGGTCTCATCTTCTTCTGGAGCTGGCAACAGGATCTCGGCTTTGTTCTTCAAGACTG 495  
QY 120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnValIysAlaTyrArgAsp-AspI 140

Search completed: November 21, 2003, 17:13:02  
Job time : 2160 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2003, 13:30:43 ; Search time 280 Seconds  
(without alignments)  
2246.318 Million cell updates/sec

Title: US-09-972-970-4

Perfect score: 1283

Sequence: 1 MPKGHQHFQBEVGGCKGYF.....RAPYTPKAWASLRSGCRTT 233

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09972970/runat\_21112003\_125335\_27199/app\_query.fasta\_1.391  
-DB=N Geneseq\_19Jun03 -QFMT=kasap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09972970@cgn 1.312 @runat\_21112003\_125335\_27199 -NCPU=6 -ICPU=3  
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	DB	ID	Description
1	1283	100.0	2538	24	AAD21883	Human TM4SF recept
2	1215.5	94.7	2091	21	AAA95482	CDNA encoding a hu
3	1215.5	94.7	3185	24	AAI72633	Human HNTWH27 codi
4	1212.5	94.5	1126	22	AAC90014	Human polynucleoti
5	1204.5	93.9	1644	24	ABL90838	Human polynucleoti
6	1176.5	91.7	1023	24	ABZ11816	Human colon cancer
7	958.5	74.7	1655	21	RAC98195	Rat gamma-hydroxyb
8	955.5	74.5	1567	23	ACC44092	Clone HE8EU16 codi
9	918.5	71.6	1174	22	AAC90015	Clone HE8EU16 codi
10	892.5	69.6	1178	22	AAC90020	Polynucleotide seq
11	807	62.9	1110	20	AAD20854	Human CDNA encodin
12	807	62.9	1110	22	AAZ59291	Human polynucleoti
13	807	62.9	1110	24	ABA90960	NET-4 antisense mo
14	790	61.6	864	24	AAI72287	CDNA sequence #126
15	759.5	59.2	1988	24	ABK35735	Human TANGO 339 CD
16	654.5	51.0	813	22	AAF90629	Human tetraspanin
17	654.5	51.0	1388	24	AAD29623	Novel human CDNA s
18	654.5	51.0	2465	25	ABX70953	Human TANGO 339 CD
19	654.5	51.0	2715	22	AAF90628	Human TANGO 339 K1
20	653.5	50.9	813	22	AAF90656	Human TANGO 339 F2
21	651.5	50.8	2623	23	AAZ92016	DNA encoding novel
22	651.5	50.8	2623	23	AAZ92016	Sequence encoding
23	650.5	50.7	813	22	AAF90658	CDNA of Human F-bo
24	650.5	50.7	813	22	AAF90659	Human colon cancer
25	641.5	50.0	2672	21	AAZ65259	Drosophila melanog
26	635.5	49.5	1428	21	AAZ93623	Human CDNA encodin
27	583.5	45.5	1667	21	AAZ65350	Human secreted pro
28	556	43.3	601	23	AAZ93664	Human tetraspanin
29	537	41.9	368	22	AAZ93364	DNA encoding novel
30	537	41.9	368	22	AAZ93364	Sequence encoding
31	490.5	38.2	839	22	AAH34927	CDNA of Human F-bo
32	482.5	37.6	810	23	ABL14835	Human colon cancer
33	476.5	37.1	852	22	AAZ44553	Drosophila melanog
34	476.5	37.1	1932	22	AAZ98047	Human CDNA encodin
35	476.5	37.1	1932	22	AAZ98047	Human secreted pro
36	476.5	37.1	1932	22	AAZ98047	Human secreted pro
37	476.5	37.1	1932	22	AAZ98047	Human secreted pro
38	461.5	36.0	2638	24	ABZ11836	Human polynucleoti
39	441	34.4	639	21	ABL90058	Human polynucleoti
40	441	34.4	2367	21	AAA60182	Hydrophobic domain
41	430	33.5	3009	23	ABL14834	Drosophila domain
42	427.5	33.3	816	21	AAZ93675	Human tetraspanin
43	421.5	32.9	1187	23	ABL07929	Drosophila melanog
44	411.5	32.1	1120	23	ABL16799	Drosophila melanog
45	402	31.3	3244	23	ABL07928	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAD21883  
ID AAD21883 standard; cDNA; 2538 BP.  
XX  
AC AAD21883;  
XX  
DT 12-FEB-2002 (first entry)

Human TM4SF receptor-encoding gene 1 cDNA clone HOF0B55, SEQ ID NO:2.  
Human; 4-transmembrane superfamily; TM4SF receptor; breast cancer;  
ovarian cancer; immune disorder; Addison's disease; wound healing;  
gene therapy; autoimmune haemolytic anaemia; autoimmune thyroiditis;  
diabetes mellitus; allergy; Crohn's disease; multiple sclerosis;  
rheumatoid arthritis; ulcerative colitis; cardiovascular disorder;  
myocardial ischaemia; neurological disease; antifungal; antiviral;  
antibacterial; cerebral anoxia; epilepsy; infectious disease;

KW antiparasitic; cancer; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 119..820

FT /\*tag= a

FT /product= "Human TM4SF receptor protein"

XX WO200177173-A1.

XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US111130.

XX 10-APR-2000; 2000US-195336P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Shi Y, Ruben SM;

XX WPI; 2002-017447/02.

XX F-PSDB; AAEI3218.

XX Novel isolated protein, a member of 4-transmembrane superfamily of  
PT receptor polypeptides, useful in the prevention, treatment and  
PT diagnosis of cancer, immune disorders, cardiovascular disorders and  
PT neurological diseases -

XX Claim 1; Page 265-266; 271pp; English.

XX The invention relates to human 4-transmembrane superfamily (TM4SF)  
CC receptor polypeptides and polynucleotides. Sequences of the invention  
CC are useful for preventing, treating, ameliorating or diagnosing a  
CC pathological condition or a susceptibility to a pathological condition.  
CC TM4SF polypeptides are useful for screening molecules which modify  
CC their activity. TM4SF nucleic acids, protein, antibodies, agonists and  
CC antagonists are useful in the diagnosis, treatment and prevention of  
CC cancer, particularly breast and ovarian cancer, and other cancers of  
CC the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,  
CC liver, lung, or urogenital; immune disorders such as Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemias; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections. TM4SF polynucleotides are  
CC also useful in gene therapy. The present sequence is human TM4SF  
CC receptor protein encoding cDNA.

XX Sequence 2538 BP; 528 A; 669 C; 747 G; 594 T; 0 other;

XX Alignment Scores:

Pred. No.: 1.59e-114 Length: 2538  
Score: 1283.00 Matches: 233  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0

US-09-972-970-4 (1-233) x AAB21883 (1-2538)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyIstYrPhe 20

Db 119 ATGCGGGGACAGCAGCAGCTTCAGGACCCGAGGTCGGCTGCTGGGAAATACTTC 178

QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40

Db 179 CTGTTGGCTTCAACATTGTTTCTGGGTCTGGGAGCCCTGTTCTGGCCATCGGCTC 238

QY 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60

Db 239 TGGGCTGGGGTGAGAGGGTGTCTCTCCACATCTCTGGCGCTGACCGATCTGGGAGGC 298

QY 61 LeuAppProValTrpLeuPheValValGlyValMetSerValLeuGlyPheAla 80  
Db 299 CTCGACCTGTGTGGCTGTGTGTAGTGGTGGAGCGTCATGTCCGTGGCTTTGCC 358  
QY 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 359 GGTGTCATCGGGCTCTCGGGAGAACACTTTCTCTGCTCAAGTGTTCAGTGTTCCTT 418  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 419 GGCCTCATCTTCTTCCTGGAGCTGGCAACAGGGATCTTGGCCCTTGGTATTCAAGGACTGG 478  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyArgAspAspIle 140  
Db 479 ATTCGAGACCACTCAATTTCTTATTAAACAACAGCTCAAGCCCTATCGGGATGACATT 538  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyTrpSerCysCysGlyAlaAArgGly 160  
Db 539 GACCTCCAGAACCTCATTTGCTCAGGAATATTGGTCTTGTCTGGGAGCCCGAGGG 598  
QY 161 ProAsnAspTrpAsnLeuAsnIleTyPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 599 CCTAATGACTGGAACCTCAATATCTATTTCAACTGCACCTGACTTGAACCCGAGCCGAGAG 658  
QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrProSer 200  
Db 659 CCCTCGGGGTGCTCTCTCTGCTGTGTGTCAGGACCCCTGCGATGTCTCAACACCAGT 718  
QY 201 ValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyThrProLys 220  
Db 719 GTGGCTATGATGTCTCGGCTCAAACTGGAGCTGGAGCAGCAGGCTCCATACACCAAG 778  
QY 221 AlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
Db 779 GCTGTGTGGCCAGTTTGAGAAGTGGCTGCAGGACAACC 817  
RESULT 2  
AAA96482  
ID AAA96482 standard; cDNA; 2091 BP.  
XX  
AC AAA96482;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE cDNA encoding a human transmembrane protein.  
KW Human; transmembrane protein; cell proliferation disorder; myeloma;  
KW reproductive disorder; smooth muscle disorder; neurological disorder;  
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;  
KW Alzheimer's disease; Tourette's disorder; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 113..1102  
FT /\*tag= a  
XX  
XX WO200056891-A2.  
XX  
XX 28-SEP-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07817.  
XX  
XX 22-MAR-1999; 99US-0125537.  
XX 16-JUN-1999; 99US-0139565.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O, Baughn MR;  
PI Lu DAM, Azimzai Y, Yang J;  
XX

DR WPI; 2000-579485/54.  
DR P-PSDB; AAB18969.  
XX  
PT New human transmembrane proteins are used to treat a disease or  
PT condition associated with decreased expression of functional HTMP e.g.  
PT Tourette's disorder, angina and leukaemia -  
XX  
XX  
PS Claim 4; Page 113-114; 130pp; English.  
XX  
CC The present sequence encodes a human transmembrane proteins (HTMP).  
CC Agonists and antagonists of the protein are used to treat a disease  
CC or condition associated with overexpression of the protein. Diseases  
CC and conditions which can be treated include cell proliferative,  
CC immunological, reproductive, smooth muscle and neurological disorders  
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency  
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,  
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The  
CC polynucleotides may be used to detect and quantify gene expression in  
CC biopsied tissues where protein expression may be correlated with disease  
CC e.g. to determine absence, presence or excess expression of HTMP or to  
CC monitor regulation of HTMP expression during therapeutic intervention.  
XX  
SQ Sequence 2091 BP; 392 A; 602 C; 637 G; 460 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 4.4e-108 Length: 2091  
Score: 1215.50 Matches: 227  
Percent Similarity: 96.80% Conservatives: 0  
Best Local Similarity: 96.80% Mismatches: 6  
Query Match: 94.74% Indels: 2  
DB: 21 Gaps: 1  
  
US-09-972-970-4 (1-233) x AAA96482 (1-2091)  
  
Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
Db 113 ATGCCCGGCAAGCAGCAGCATTTCCAGGAACCTCAGGTCGGCTCGCGGAAATACTTC 172  
  
Qy 21 LeuPheGlyPheAsnIleValPheTyrValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 173 CTGTTTGGCTTCAACATGTCTTCTGGGTGCTGGAGCCCTTCCTCGCTATCGGCTC 232  
  
Qy 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 233 TGGGCTTGGGTGAGAGGGGCTTCTCTCGAACATCTCAGCGCTGACAGATCTGGAGGC 292  
  
Qy 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 293 CTTGACCCCGTGGCTGTTTGTGTGTGTGTGGAGCGTCATGCTGCTGGCTTTGCT 352  
  
Qy 81 GlyCysIleGlyAlaLeuAlaGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 353 GGCTGCAATGGGGCCCTCCGGAGAACACCTTCTGCTCAAGTGTTCCTCGTGTCTCCTC 412  
  
Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 413 GGTCTCATCTTCTCTGGAGCTGGCAACGGATCTGGCTTTGTCTTCAAGACTGG 472  
  
Qy 121 IleArgAspGlnIleuAsnPhePheIleAsnAsnValLysAlaIleTyrArgAspAspIle 140  
Db 473 ATTGAGACCACTCAACCTTCTTATCAACAACAACGCTCAAGGCTTACCGGAGCAGATT 532  
  
Qy 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyValAlaArgGly 160  
Db 533 GACCTCCAGAACCTCATGTGCTTGTCTCAGGAATACGTGCTTCTCGGAGCCCGAGGC 592  
  
Qy 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 593 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCGAGCGGAG 652  
  
Qy 181 ArgCysGlyValProPheSerCysCysValArgAspProIle----MetSerSerThrPr 199  
Db 653 CGCTGGGGGTGCTTCTCTGCTGCTCAGGAGACCTCGGAGGATGTCTCTCAACACC 712

Qy 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219  
Db 713 CAGTGTGGCTACGACGCTCCGGCTCAAACTGGAGCTGGAGCGGCTTCAATCCACACC 772  
  
Qy 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
Db 773 AAAGGCTCGTGGGCCAGTTTGAGAGCTGGCTGCAGGACACC 815  
  
RESULT 3  
AAI72633  
ID AAI72633 standard; cDNA; 3185 BP.  
AC AAI72633;  
XX  
XX 18-JUN-2002 (first entry)  
Human 23228 cDNA.  
Gene; tetraspanin; 23228; cell surface protein; transmembrane domain;  
extracellular growth factor; HB-EGF; TGF-alpha; amphiregulin;  
diabetes mellitus; arthritis; multiple sclerosis; encephalomyelitis;  
dermatitis; Crohn's disease; and asthma; cancer; metastasis;  
viral infection; cellular signalling activity; cell proliferation;  
cell motility; CD81; B-Cell antigen receptor; ss.  
Homo sapiens.  
OS Homo sapiens.  
FH  
FT Key Location/Qualifiers  
FT CDS 168..980  
FT /\*tag= a  
FT /product= "23228"  
XX  
XX WO200216603-A2.  
XX  
XX 28-FEB-2002.  
XX  
XX 21-AUG-2001; 2001WO-US41811.  
XX  
XX 21-AUG-2000; 2000US-226612P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Leiby KR;  
XX  
XX WPI; 2002-315466/35.  
XX P-PSDB; AAB47950.  
XX  
XX New tetraspanin 23228 polypeptide useful in screening assays,  
predictive medicine and as a prophylactic or therapeutic agent, e.g.,  
for hematopoietic and immune diseases such as diabetes or multiple  
sclerosis -  
XX  
XX Claim 1; Page 111-113; 123pp; English.  
XX  
XX This sequence encodes tetraspanin 23228. This protein is a cell surface  
protein having four transmembrane domains (TM1-TM4). TM1, TM3 and TM4  
have a single polar amino acid located within them, which may interact  
with each other and contribute to domain stability. The cytoplasmic N-  
and C-terminal domains and the intracellular loop between TM2 and TM3  
are less than 30 amino acids in length. The second extracellular domain  
between TM3 and TM4 contains conserved Cys residues and may function  
to bind extracellular growth factors, such as HB-EGF, TGF-alpha and  
amphiregulin. The 23228 protein is useful for diagnosing and treating  
23228-mediated disorders, e.g., haematopoietic and/or immune disorders  
such as diabetes mellitus, arthritis, multiple sclerosis,  
encephalomyelitis, dermatitis, Crohn's disease and asthma; cell  
proliferation and differentiation disorders, e.g., cancers or  
metastasis; and/or viral infections. 23228 polypeptide can control  
cellular signalling activity, bind to an extracellular growth factor,  
for example, amphiregulin, regulate cell proliferation, bind to a cell  
surface protein, to recruit intracellular kinases, to regulate cell  
motility, bind to another tetraspanin such as CD81, to associate with a



CC B-Cell antigen receptor and the ability to modulate the association  
 CC with a virus.  
 XX Sequence 3185 BP; 676 A; 883 C; 945 G; 681 T; 0 other;

SQ Alignment Scores:

Pred. No.:	7.7e-108	Length:	3185
Score:	1215.50	Matches:	227
Percent Similarity:	96.60%	Conservative:	0
Best Local Similarity:	96.60%	Mismatches:	6
Query Match:	94.74%	Indels:	2
DB:	24	Gaps:	1

US-09-972-970-4 (1-233) x AAI72633 (1-3185)

Qy	1	MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe	20
Db	168	ATGCCCGGCAAGCACCAGCATTTCCAGAACCTTGAGGTGGCTGCTGGCGGAATACTTC	227
Qy	21	LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu	40
Db	228	CTGTTGGCTTCAACATTTGCTTCTGGGTGGAGCCCTGTTCTGGCTATCGGCCTC	287
Qy	41	TrpAlaTrpGlyGlyLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly	60
Db	288	TGGGCTGGGTGAGAGGGCTTCTTCGAACATCTCAGCGCTGACAGATCTGGGAGGC	347
Qy	61	LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla	80
Db	348	CTTGACCCGCTGGCTGTTGTTGGTAGTGGAGCGTCATGTCGGTCTGGCTTTCCT	407
Qy	81	GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPheSerValPheLeu	100
Db	408	GGCTGCATGGGGCCCTCGGAGAACACCTTCTGCTCAAGTTTTCCTGGTTCCTC	467
Qy	101	GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp	120
Db	468	GGTCTCATCTTCTTCTGGAGTGGCAACAGGATCTCTGGCTTGTCTTCAAGGACTGG	527
Qy	121	IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle	140
Db	528	ATTGAGACCAAGCTCAACCTTTCATCAACAAACAGTCAAGGCTACCGGAGCAGCAT	587
Qy	141	AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly	160
Db	588	GACCTCCAGAACCTCATTTGCTCAGGAATACTGCTTCTGCTGGAGCCGAGGC	647
Qy	161	ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu	180
Db	648	CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCGGGAG	707
Qy	181	ArgCysGlyValProPheSerCysCysValArgAspProAla---MetSerSerThrPr	199
Db	708	CGCTGGGGTGCCCTTCTCTGCTGCGTCAAGGACCTCGGAGGATGCTCTCAACACC	767
Qy	199	oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrTrp	219
Db	768	CAGTGTGGCTACGACGTCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTTTCATCCAC	827
Qy	219	OlyAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr	233
Db	828	AAAGGCTGCGTGGGCCAGTTTGGAGAAGTGGCTGCAGGAGCAACC	870

RESULT 4

AAC90014

ID AAC90014 standard; cDNA; 1126 BP.

XX AAC90014;

XX AAC90014;

AC 09-MAR-2001 (first entry)

DT Clone HNTMH27 coding sequence.

XX

XX

XX

XX

XX

XX

XX

XX

KW Gene therapy; human; 4 transmembrane superfamily receptor protein;  
 KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
 KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
 KW bacterial infection; viral; fungal; ss.  
 XX Homo sapiens.  
 OS WO200070076-A1.  
 PN 23-NOV-2000.  
 PD 18-MAY-2000; 2000WO-US13504.  
 PF 19-MAY-1999; 99US-0135122.  
 PR 03-JUN-1999; 99US-0137797.  
 PR 11-JUN-1999; 99US-0138573.  
 PR 18-AUG-1999; 99US-0149447.  
 PR 28-JAN-2000; 2000US-0178770.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;  
 XX Rosen CA;  
 PI WPI: 2001-007502/01.  
 XX P-PSDB; AAB49504.  
 DR Isolated nucleic acid molecule encoding human soluble 4 transmembrane  
 XX superfamily receptor protein, useful for diagnosing, treating and/or  
 XX preventing disorders e.g. Alzheimer's, cancer and arrhythmia -  
 PS Claim 1; Pages 269-270; 297pp; English.  
 CC The present invention relates to isolated nucleic acids and proteins  
 CC encoding human soluble 4 transmembrane superfamily receptor protein (see  
 CC AAC90012-C90023 and AAB49502-B49513). The present sequence is one such  
 CC nucleic acid. The present sequence is useful for preventing, treating or  
 CC ameliorating a medical condition and in diagnosing (susceptibility to) a  
 CC pathological condition e.g. endocrine disorders e.g. Addison's disease,  
 CC (cardio)vascular diseases e.g. arrhythmia and atherosclerosis, and  
 CC cerebrovascular diseases, neural disorders e.g. Alzheimer's and  
 CC Parkinson's disease, reproductive disorders, skin disorders e.g.  
 CC psoriasis, renal system disorders e.g. nephritis, (auto)immune system  
 CC disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.  
 CC neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections  
 CC caused by bacteria, viruses and fungi.  
 XX SQ Sequence 1126 BP; 191 A; 352 C; 343 G; 240 T; 0 other;

Alignment Scores:

Pred. No.:	3.77e-108	Length:	1126
Score:	1212.50 <td>Matches:</td> <td>226</td>	Matches:	226
Percent Similarity:	96.60%	Conservative:	1
Best Local Similarity:	96.17%	Mismatches:	6
Query Match:	94.51%	Indels:	2
DB:	22	Gaps:	1

US-09-972-970-4 (1-233) x AAC90014 (1-1126)

Qy	1	MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe	20
Db	157	ATGCCCGGCAAGCACCAGCATTTCCAGAACCTTGAGGTGGCTGCTGGCGGAATACTTC	216
Qy	21	LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu	40
Db	217	CTGTTGGCTTCAACATTTGCTTCTGGGTGGAGCCCTGTTCTGGCTATCGGCCTC	276
Qy	41	TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly	60
Db	277	TGGGCTGGGGTGAGAGGGCGGTCTCTCGAACATCTCAGCGCTGACAGATCTCGGAGGC	336
Qy	61	LeuAspProValTrpLeuPheValValGlyValMetSerValLeuGlyPheAla	80



Db 337 CTTGACCCCGTGTGGCTGTTTGTGGTAGTGGAGCGTCATGTGCGTGTGGGCTTTGCT 396  
 QY 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPheSerValPheLeu 100  
 Db 397 GGTGTCATTTGGGCGCTCCGGGAGACACCTCTCTCAAGTPTTTTCTCCGTTCTCTC 456  
 QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTTP 120  
 Db 457 GGTCTCATCTTCTCTGGAGCTGGCAACAGGATCTCTGGCTTTGTCTTCAAGGACTGG 516  
 QY 121 IleArgAspGlnLeuAenPhePheIleAenAsnValLysAlaTyrArgAspAspIle 140  
 Db 517 ATTGAGACGAGCTCAACCTCTTCAACAACAACAGCTCAAGGCTTACCGGACGACATT 576  
 QY 141 AspLeuGlnAenLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160  
 Db 577 GACCTCCAGAACCTCAATGACTTTGCTCAGGAATACTGGTCTTGTGGAGCCGAGGC 636  
 QY 161 ProAsnAspTrpAsnLeuAenIleTyrPheAenCysThrAspLeuAenProSerArgGlu 180  
 Db 637 CCNATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCGGAG 696  
 QY 181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199  
 Db 697 CGCTGCGGGTGCCTTCTCTCTGCTCGTCAGGACCTCGGAGGATGTCTCAACACC 756  
 QY 199 oSerValAlaMetMetSerGlySerAenThrTrpSerSerArgAlaProTyrThrPr 219  
 Db 757 CAGTGTGGCTAGACGCTCGGCTCAAACTGGAGCTGGAGCAGCAGGCGCTTCATCCACACC 816  
 QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
 Db 817 AAAGGCTGCGTGGCGCAGTTTGAGAAATGGCTGCAGGACAACC 859  
 RESULT 5  
 ABL90838  
 ID ABL90838 standard; cDNA; 1644 BP.  
 AC ABL90838;  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 1400.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US16450.  
 XX  
 PR 19-MAY-2000; 2000US-205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-122018/16.  
 XX  
 PT P-PSDB; ABL90429.  
 XX  
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 PS Claim 4; SEQ ID NO 1400; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL99449-ABL90853) and proteins  
 CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1644 BP; 297 A; 479 C; 513 G; 346 T; 9 other;

Alignment Scores:  
 Pred. No.: 3,73e-107 Length: 1644  
 Score: 1204.50 Matches: 225  
 Percent Similarity: 95.74% Conservative: 0  
 Best Local Similarity: 95.74% Mismatches: 8  
 Query Match: 93.88% Indels: 2  
 DB: 24 Gaps: 1

US-09-972-970-4 (1-233) x ABL90838 (1-1644)  
 QY 1 MetProGlyLYeHieGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 Db 222 ATGCCCGGAGCAGCAGCAGCATTTCCAGGAACCTGAGTGGCTCGGCTCGCGGAATACTTC 281  
 QY 21 LeuPheGlyPheAenIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 282 CTGTTGGCTTCAACATTTCTCTGGGTGCTGGGAGCCCTGTCTCTGGCTATCGGCTC 341  
 QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAenIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 342 TGGGCTGGGTGAGAGGGCGTTCTCGAACATCTYAGCGCTGACATCTGGGAGGC 401  
 QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
 Db 402 CTTGAMCCGCTGGCTGTTTGTGTAGTTGGAGGCGTCATGTCGGTGTGGGCTTTGCT 461  
 QY 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPheSerValPheLeu 100  
 Db 462 GGTGCTATGGGGCCCTCGGAGAACACCTTCTCTCAAGTTTTCTCCGTTTCTCTC 521  
 QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTTP 120  
 Db 522 GGTCTCATCTTCTCTGGAGCTGGCAACAGGATCTCTGGCTTTGTCTTCAAGGACTGG 581  
 QY 121 IleArgAspGlnLeuAenPhePheIleAenAenValLysAlaTyrArgAspAspIle 140  
 Db 582 ATTCGAGACGAGCTCAACCTCTTCAACAACAACAGTCAAGGCTTACCGGACGACATT 641  
 QY 141 AspLeuGlnAenLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160  
 Db 642 GACCTCCAGAACCTCAATGACTTTGCTCAGGAATACTGGTCTTGTGGAGCCGAGGC 701  
 QY 161 ProAenAspTrpAsnLeuAenIleTyrPheAenCysThrAspLeuAenProSerArgGlu 180  
 Db 702 CCCAATGACTGGAACCTCAATATCTACTTCAACTGCACTGACTTGAACCCCGGAGGC 761  
 QY 181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199  
 Db 762 CGCTGCGGGTGCCTTCTCTCTGCTGCTCAGGAGCCCTCGGAGGATGTCTCTCAACACC 821  
 QY 199 oSerValAlaMetMetSerGlySerAenThrTrpSerSerArgAlaProTyrThrPr 219

Db 822 CAGTGTGGCTACGAGCTCGCGCTCAAACTGGAGCTGGAGCAGGAGCTTCATCCACACC 881  
 QY 219 OlysisAlaValTrrPALaserLeuArgSerGlyCysArgThr 233  
 Db 882 AAAGGCTGGCTGGCGCCAGTTTGAGAGTGCTGCAGGACAACC 924

## RESULT 6

ABZ11816

ID ABZ11816 standard; cDNA; 1023 BP.

XX AC ABZ11816;  
 XX 20-JAN-2003 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 698.  
 XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 KW anti-parkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic; gene; ss.

Homo sapiens.

OS XX  
 XX XX  
 PN WO200270539-A2.

PD 12-SEP-2002.

XX PF 05-MAR-2002; 2002WO-US05095.

XX PR 05-MAR-2001; 2001US-0799451.

XX PA (HYSE-) HYSEQ INC.

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao OA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2002-759812/82.

DR P-PSDB; ABP69599.

XX New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
 PT platelet or coagulation disorders

XX Claim 1; SEQ ID NO 698; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (1) comprising a  
 CC nucleotide sequence selected from any of 948 sequences  
 CC (ABZ1119-ABZ12066) or their mature protein coding portion, active domain  
 CC coding protein or complementary sequences. The polynucleotides are useful  
 CC for identifying expressed genes or for physical mapping of human genome.  
 CC The encoded polypeptides (ABP69902-ABP69849) are useful as molecular  
 CC weight markers, as a food supplement, for generating antibodies, in  
 CC medical imaging, screening and diagnostic assays and for treating  
 CC cell-proliferative disorders (cancer), neurodegenerative diseases  
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
 CC disorders, platelet or coagulation disorders, wound, burns, incision,  
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
 CC parasitic), arthritis, etc.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1023 BP; 190 A; 300 C; 293 G; 240 T; 0 other;  
 SQ

Alignment Scores:

Pred. No.: 1,04e-104 Length: 1023  
 Score: 1176.50 Matches: 227  
 Percent Similarity: 95.38% Conservative: 0  
 Best Local Similarity: 95.38% Mismatches: 6  
 Query Match: 91.70% Indels: 5  
 DB: 24 Gaps: 1

US-09-972-970-4 (1-233) x ABZ11816 (1-1023)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 Db 31 ATGCCCGCAAGCACCAGCATTTCCAGGAACCTGAGGTGGCTGCGCGAAATACTTC 90

QY 21 LeuPheGlyPheAsnIleValPheTrrValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 91 CTGTTTGGCTTCAACATTGTCTTCTGGGTGGAGCCCTGTCTCTGGGTATCGGGCTC 150

QY 41 TrrAlaTrrGlyCylLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 151 TGGGCTGGGGTGAGNAGGGCGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 210

QY 61 LeuAspProValTrr-LeuPheValValValGlyGlyValMetSerValLeuGlyPheAl 80  
 Db 211 CTTGACCCCGTGTGGCTTCTTGTGTGTAGTTGGAGCGTCATGTCGTGCTGGGCTTTC 270

QY 80 a-GlyCys-IleGlyAlaLeuArgGluAenThrPheLeuLeuLysPheSerValPhe 99  
 Db 271 TGGGCTTGCAATTGGGGCCCTCCGGGAGAACACCTTCTCTCAAGTTTTTCTCCGTGTTC 330

QY 100 LeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAsp 119  
 Db 331 CTGGGTCTCATCTTCTTCTGGAGCTGGGCAACAGGATCTTGGCCCTTGTCTTCAAGGAC 390

QY 120 TrrIleArgAspGlnLeuAenPheIleAenAsnValLysAlaTyrArgAspAsp 139  
 Db 391 TGGATTTCAGAGCAGCTCAACCTTTCATCAACAACAGCTCAAGGCTACCGGAGCGAC 450

QY 140 IleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrrSerCysGlyAlaArg 159  
 Db 451 ATTGACCTCCAGAACCTCATTGCTTGTCTAGGAATACCTGGTCTTGTCTGGGAGGCCGA 510

QY 160 GlyProAsnAspTrrAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArg 179  
 Db 511 GGCCCCAATGACTGGAACTCAATATCTACTTCACTGCTGCTGAAACCCAGCCCG 570

QY 180 GluArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerTh 198  
 Db 571 GAGCGCTGGCGGGTGCCCTTCTCTGCTGCGTCAAGGACCTTGGCGAGGATGTCTCAAC 630

QY 198 rProSerValAlaMetMetSerGlySerAenThrPrrSerSerArgAlaProTyrTh 218  
 Db 631 ACCCAGTGTGGCTACGACGTCCGGCTCAAACTGGAGCTGGAGCAGGAGGCTTCATCCAC 690

QY 218 rProLysAlaValTrrPalaserLeuArgSerGlyCysArgThrThr 233  
 Db 691 ACCAAGGCTGGCTGGCGCCAGTTTTCAGAAAGTGCTGCGAGGACCAACC 736

## RESULT 7

AAC98195

ID AAC98195 standard; cDNA; 1655 BP.

XX AAC98195;

XX AC AAC98195;

XX 09-MAR-2001 (first entry)

XX Human colon cancer antigen nucleotide sequence SEQ ID NO:205.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW identification; cytostatic; cardioactive; neuroprotective; vulnery;  
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 KW nephrotropic; antinfertive; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;

KW infectious disease; cardiovascular disorder; ss.

OS Homo sapiens.

XX WO20005351-A1.

PN 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587534/55.

DR P-PSDB; AAB53438.

XX Colon cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer -

PS Claim 1; Page 628-629; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
CC human colon cancer antigens can have cytostatic, cardioactive, muscular,  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
CC vulnary, nephrotropic, antineoplastic and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridisation probes. The proteins  
CC may also be used to prevent diseases such as neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, wounds, renal disorders, infectious  
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
CC invention.

XX SQ Sequence 1655 BP; 371 A; 415 C; 437 G; 425 T; 7 other;

Alignment Scores:

Pred. No.: 2,87e-83 Length: 1655  
Score: 958.50 Matches: 180  
Percent Similarity: 84.19% Conservative: 17  
Best Local Similarity: 76.92% Mismatches: 33  
Query Match: 74.71% Indels: 4  
DB: 21 Gaps: 2

US-09-972-970-4 (1-233) x AAC98195 (1-1655)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTrpPhe 20  
Db 277 ATGTCGGGAG-----CACTACAGGGTCTCGAAGTCAGTTGTCATCAAACTCTC 330  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 331 ATATTGGCTTCAATGTGCATATTTTGGTTTGGGAATAACATTTCTTGGAAATTGACTG 390  
QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 391 TGGGCATGGAATGAAAGAGGATTTGTCCACATCTCTTCCATCACCAGTCTCGGCGGC 450  
QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 451 TTTGACCCAGTTGGCTCTTCTTGTGGTGGGAGGAGTGATGTTTCATTTGGGATTGCA 510  
QY 81 GlyCysIleGlyAlaLeuArgGluAnThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 511 GGGTCATTTGGAGCGCTACGGGAAACACTTTCTCTCAAGTTTTTTCTGTGTCCTG 570

QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 571 GGAATTATTTCCTCTGGAGCTCACTGCCGGAGTTCTAGCATTTGTTTCAAGACTGG 630  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaIleValArgAspAspIle 140  
Db 631 ATCAAGAGACCGCTGATTTCCTTTATTAACAACAACATCAGAGCATATCGGATGACATT 690  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysGlyAlaArgGly 160  
Db 691 GATTTGCAAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGTCTGGGGCTTTTGA 750  
QY 161 ProAsnAspTrpAsnLeuAsnIleTrpPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 751 GCTGATGATTGGAAACCTAATATTCTCAATTGCACAGATTCCAATGCAAGTCGAGAG 810  
QY 181 ArgCysGlyValProPheSerCysValArgAspProAla---MetSerSerThrPr 199  
Db 811 CGATGTGGCGTTCATTCTCTGCTGCTAAAGATCCCGCAGAAGATGTCATCAACT 870  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTrpThrPr 219  
Db 871 CAGTGTGGCTATGATGCCAGGCAAAACCAAGTTGACCAGCAGATTGTAATCTACAG 930  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
Db 931 AAAGGCTGTGTGCCCGAGTTTGAGAAAGTGGTTGCAGGACA 970

RESULT 8

ACC44092

ID ACC44092 standard; DNA; 1567 BP.

XX ACC44092;

XX 13-JUN-2003 (first entry)

DE Rat gamma-hydroxybutyrate receptor cDNA.

KW gene; rat; gamma-hydroxybutyrate; anti-epileptic; anxiolytic; ds;  
KW antineurodegeneration; antipsychotic; brain; dopamine; opioid; GABA;  
KW gamma-aminobutyric acid; diagnosis; epilepsy; anxiety; sleep disorder;  
KW behavioral disorder; neurodegeneration; Parkinson's disease; psychosis;  
KW schizophrenia.

OS Rattus rattus.

XX Key Location/Qualifiers

XX CDS 22..1560

XX FT /tag= a

XX FT /product= "gamma-hydroxybutyrate receptor"

XX WO200078948-A2.

XX 28-DEC-2000.

XX 19-JUN-2000; 2000WO-PR01687.

XX 18-JUN-1999; 99FR-0007784.

XX (UYPA-) UNIV PASTEUR LOUIS.

XX Andriananpandry C, Maitre M;

XX WPI; 2001-091570/10.

XX P-PSDB; ABP98695.

XX New rat brain gamma-hydroxybutyrate receptor and its encoding nucleic  
XX acid, useful for identifying agents for treating e.g. epilepsy -  
XX Claim 5; Page 64-66; 66pp; French.  
XX This sequence represents the cDNA sequence encoding a novel rat

CC gamma-hydroxybutyrate receptor (GHR) isolated from rat brain hypocampal  
 CC cell cDNA library. The invention relates to the isolation of this novel  
 CC sequence, fragments of it and homologues of the sequence except for those  
 CC homologues that having GenBank accession numbers AAC 17120 (human  
 CC tetraepan NET-4), AA615405, AA967250 or A1467230. GHR is the receptor  
 CC for gamma-hydroxybutyrate in the rat brain and is involved in regulating  
 CC dopaminergic, opioid and GABA(gamma-aminobutyric acid)ergic activities.  
 CC The nucleic acid that encodes the protein is used: (1) as primers or  
 CC probes for detection/amplification, particularly for screening gene  
 CC libraries to identify promoters and regulators of the GHR gene; (2) for  
 CC expression of recombinant polypeptides; and (3) to detect allelic  
 CC variants, mutations, deletions, loss of heterozygosity or genetic  
 CC abnormalities in the GHR gene (for diagnosing diseases, or  
 CC susceptibility, associated with abnormal expression of GHR). The  
 CC protein for cells and transgenic animals expressing it) is used:  
 CC (1) to screen for agents that interact with GHR; (2) to study  
 CC expression/activity of the receptor, including its interaction with  
 CC other compounds; and (3) to raise antibodies (Ab) specific for GHR.  
 CC The Ab are used to detect/measure (1) in diagnostic immunoassays.  
 CC The agents are used to prevent or treat diseases associated with abnormal  
 CC expression or activity of GHR, particularly those involving cerebral  
 CC GABA(gamma-aminobutyric acid)ergic and/or dopaminergic activities, e.g.  
 CC epilepsy, anxiety, sleep or behavioral disorders, withdrawal from  
 CC addictive drugs, neurodegeneration (Parkinson's disease), psychoses,  
 CC schizophrenia and regulation of secretion of hormones (growth hormone  
 CC and prolactin) that are under dopaminergic control.

SQ Sequence 1567 BP; 281 A; 436 C; 480 G; 370 T; 0 other;

Alignment Scores:  
 Pred. No.: 5,22e-83 Length: 1567  
 Score: 955.50 Matches: 193  
 Percent Similarity: 85.11% Conservatives: 7  
 Best Local Similarity: 82.13% Mismatches: 29  
 Query Match: 74.47% Indels: 8  
 DB: 23 Gaps: 1

US-09-972-970-4 (1-233) x ACC44092 (1-1567)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 DB 135 ATCCCGCGCAAGCATCAGCAATTCAGGACCCCTGAGCTGGCTCGCGGAAATACTTT 194  
 QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyValAlaLeuPheLeuAlaIleGlyLeu 40  
 DB 195 CTGTTTGGCTTCAACATTGTCTTCTGGTCTGGAGCCCTGTCTCTGGCCATCGGCCTC 254  
 QY 41 TrpAlaTrpGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAsp-LeuGlyGI 60  
 DB 255 TGGGCTGGGCTCAGAAAGGCGTTCCTTCCACATCTCAGGCGGACAGATCCAGGCGG 314  
 QY 60 YLeuAppProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAl 80  
 DB 315 TCTTGACCCCGTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 374  
 QY 80 aGlyCysIleGlyValAlaLeuArgGluAsnThrPheLeuLeuValPhePheSerValPheLe 100  
 DB 375 CGGCTGCATTGGGCGCTCCGGGAAACACCTTCTCTGCTCAAAATTTCTCTGTGTCT 434  
 QY 100 uGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120  
 DB 435 CGGCTCATCTTCTCTGGAGCTGGCGCGGATCTTGGCTTCTGTGTTCAGGATTG 494  
 QY 120 pIleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspTrl 140  
 DB 495 GATCCGAGACCACTTAACCTTCTTCATCAACAACAATGTCAAGCCCTACCGGACCATAT 554  
 QY 140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyValAlaArgGI 160  
 DB 555 TGACCTTCAGACCTTATCGACTTCTCAGGATTAAGTCTGTGTGTGTGTGTGTGTGTGT 614  
 QY 160 YProAsnAspTrpAsnLeuAsnIleTyr--PheAsnCysThrAspLeu-AsnProSerAr 179

Db 615 GCCCAATGACTGGAACTCAACATCCGAGACTTCAACTCGCACTTCAAAACCCAGCCG 674  
 QY 179 gGluArgCysGlyValProPheSerCysCysValArgAspProAlaMetSerSerThrPr 199  
 Db 675 CGAGCGCTGTGGGGTGCCTTCTTCTGCTGGGTAAG-GACCTCGCGAAGACGTCCTCAAT 733  
 QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219  
 Db 734 ACCCATGTGCTATATACATCGCTCAAACTG-AGCTGGAGCACAAGATTCTATACACAA 792  
 QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArg 231  
 Db 793 AGCTGGT-----GGCATTTGAGAAAGTGGCTCAAGA 823

# RESULT 9

AAC90015  
 ID AAC90015 standard; cDNA; 1174 BP.  
 XX AAC90015;  
 AC AAC90015;  
 DT 09-MAR-2001 (first entry)  
 XX  
 XX Clone HE8EJ16 coding sequence #1.  
 DE  
 DE Gene therapy; human; 4 transmembrane superfamily receptor protein;  
 KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
 KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
 KW bacterial infection; viral; fungal; ss.  
 XX Homo sapiens.  
 XX  
 XX WO200070076-A1.  
 XX  
 XX 23-NOV-2000.  
 XX  
 XX 18-MAY-2000; 2000WO-US13504.  
 XX  
 XX 19-MAY-1999; 99US-0135122.  
 PR 03-JUN-1999; 99US-0137797.  
 PR 11-JUN-1999; 99US-0138573.  
 PR 18-AUG-1999; 99US-0149447.  
 PR 28-JAN-2000; 2000US-0178770.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;  
 PI Rosen CA;  
 XX  
 XX WPI; 2001-007502/01.  
 DR P-PSDB; AAB49505.  
 DR  
 XX  
 XX Isolated nucleic acid molecule encoding human soluble 4 transmembrane  
 PT superfamily receptor protein, useful for diagnosing, treating and/or  
 PT preventing disorders e.g. Alzheimer's, cancer and arrhythmia -  
 XX  
 XX Claim 1; Page 270; 297pp; English.  
 PS  
 XX The present invention relates to isolated nucleic acids and proteins  
 CC encoding human soluble 4 transmembrane superfamily receptor protein (see  
 CC AAC90012-C90023 and AAB49502-B49513). The present sequence is one such  
 CC nucleic acid. The present sequence is useful for preventing, treating or  
 CC ameliorating a medical condition and in diagnosing (susceptibility to) a  
 CC pathological condition e.g. endocrine disorders e.g. Addison's disease,  
 CC (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,  
 CC cerebrovascular diseases, neural disorders e.g. Alzheimer's and  
 CC Parkinson's disease, reproductive disorders e.g. skin disorders e.g.  
 CC psoriasis, renal system disorders e.g. nephritis, (auto)immune system  
 CC disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.  
 CC neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections  
 CC caused by bacteria, viruses and fungi.  
 XX  
 XX Sequence 1174 BP; 306 A; 259 C; 291 G; 316 T; 2 other;

Alignment Scores:  
 Pred. No.: 1,39e-79 Length: 1174  
 Score: 918.50 Matches: 179  
 Percent Similarity: 83.40% Conservative: 17  
 Best Local Similarity: 76.17% Mismatches: 34  
 Query Match: 71.59% Indels: 6  
 DB: 22 Gaps: 2

US-09-972-970-4 (1-233) x AAC90015 (1-1174)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 |||:|||||  
 Db 93 ATGTCCGGGAAG-----CACTACAAAGGGTCTGGAAGTCAGTTGTTGCATCAAAATCTTC 146

QY 21 LeuPheGlyPheAsnIleValTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 |||:|||||  
 Db 147 ATATTTGGCTTCAATGTCATATTTTGGTTTGGGAATAACATTTCTTGGATTTGACTG 206

QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 |||:|||||  
 Db 207 TGGCATGGAATCAAAAGAGGTTCTGTCTCAACATCTCTTCCATCACCGATCTCGGCGC 266

QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
 |||:|||||  
 Db 267 TTGACCCAGTTGGCTCTCTTGTGTGGTGGAGGAGTCATGTTCAATTTGGGATTTGCA 326

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
 |||:|||||  
 Db 327 GGGTGATTTGGAGCGTA-CGGGAAACACTTTCTTCTCAAGTTTTTTCTGTGTTCCTG 385

QY 101 GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120  
 |||:|||||  
 Db 386 GGAATTTATTTCTTCTCGAGCTCACTCGCGAGTCTTAGCATTTGTTTCAAGACTG 445

QY 120 PileArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyArgAspIle 140  
 |||:|||||  
 Db 446 GATCAAGACCACTGTATTTCTTTATAAACACACATCAGAGCATATCGGATGCAT 505

QY 140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgG 160  
 |||:|||||  
 Db 506 TGATTTGCAAAACCTCATAGATTCCACCAGGAATATTGGCAGTCTGTGGGCTTTTGG 565

QY 160 yProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArg 180  
 |||:|||||  
 Db 566 AGCTGATGATTGGAACCTAAATATTACTTCAATTGCACAGATTCCAATGCAAGTCGAGA 625

QY 180 uArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrP 199  
 |||:|||||  
 Db 626 GCGATGTGGCGTTCATTTCTCTGCTGCATAAAGATCCCGCAGAGATGTCATCAACAC 685

QY 199 roSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrP 219  
 |||:|||||  
 Db 686 TCAGTGTGGCTATGATGCCAGCAAAACCAAGTTGTACCGAGATTGTAATCTACAC 745

QY 219 roLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
 |||:|||||  
 Db 746 GAAAGCTGTGTGCCCCAGTTTGAGAAGTGGTTGCAGGACA 786

RESULT 10  
 AAC90020  
 ID AAC90020 standard; cDNA; 1178 BP.  
 XX  
 AC AAC90020;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Clone HE8EL16 coding sequence #2.  
 XX  
 KW Gene therapy; human; 4 transmembrane superfamily receptor protein;  
 KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
 KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
 KW bacterial infection; viral; fungal; ss.

OS Homo sapiens.  
 XX  
 PN WO200070076-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 18-MAY-2000; 2000WO-US13504.  
 XX  
 PR 19-MAY-1999; 99US-0135122.  
 PR 03-JUN-1999; 99US-0137797.  
 PR 11-JUN-1999; 99US-0138573.  
 PR 18-AUG-1999; 99US-0149447.  
 PR 28-JAN-2000; 2000US-0178770.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;  
 PI Rosen CA;  
 XX  
 WI: 2001-007502/01.  
 DR P-PSDB; AAB49510.  
 XX  
 PT Isolated nucleic acid molecule encoding human soluble 4 transmembrane  
 PT superfamily receptor protein, useful for diagnosing, treating and/or  
 PT preventing disorders e.g. Alzheimer's, cancer and arrhythmia -  
 XX  
 PS Claim 1; Pages 274-275; 297pp; English.  
 XX  
 CC The present invention relates to isolated nucleic acids and proteins  
 CC encoding human soluble 4 transmembrane superfamily receptor protein (see  
 CC AAC90012-C90023 and AAB49502-B49513). The present sequence is one such  
 CC nucleic acid. The present sequence is useful for preventing, treating or  
 CC ameliorating a medical condition and in diagnosing (susceptibility to) a  
 CC pathological condition e.g. endocrine disorders e.g. Addison's disease,  
 CC (cardio)vascular diseases e.g. arrhythmia and atherosclerosis,  
 CC Parkinson's disease, reproductive disorders, skin disorders e.g.  
 CC psoriasis, renal system disorders e.g. nephritis, (auto)immune system  
 CC disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.  
 CC neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections  
 CC caused by bacteria, viruses and fungi.  
 XX  
 SQ Sequence 1178 BP; 310 A; 261 C; 291 G; 316 T; 0 other;

Alignment Scores:  
 Pred. No.: 4.67e-77 Length: 1178  
 Score: 892.50 Matches: 179  
 Percent Similarity: 82.70% Conservative: 17  
 Best Local Similarity: 75.53% Mismatches: 34  
 Query Match: 69.56% Indels: 8  
 DB: 22 Gaps: 2

US-09-972-970-4 (1-233) x AAC90020 (1-1178)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 |||:|||||  
 Db 91 ATGTCCGGGAAG-----CACTACAAAGGGTCTGGAAGTCAGTTGTTGCATCAAAATCTTC 144

QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 |||:|||||  
 Db 145 ATATTTGGCTTCAATGTCATATTTGTTTGGGAATAACATTTCTTGGATTTGACTG 204

QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 |||:|||||  
 Db 205 TGGCATGGAATCAAAAGAGGTTCTGTCTCAACATCTCTTCCATCACCGATCTCGGCGC 264

QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
 |||:|||||  
 Db 265 TTGACCCAGTTGGCTCTCTTGTGTGGAGGAGTGTGATTTCAATTTGGGATTTGCA 324

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSer-ValPheLeu 100  
 |||:|||||  
 Db 325 GGGTGATTTGGAGCGTA-CGGGAAACACTTTCTTCTCAAGTTTTTTCTCGTGTCTCT 383

Qy	100	u-GlyLeuIlePhe-PheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheIysAsp	119
Db	384	CGGGAAATTATTTCTTCTCGAGCTCACTGCCGGAGTTCAGCATTTGTTTCAAAGAC	443
Qy	120	TrpIleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrrArgAspAsp	139
Db	444	TGSATCAAAGACCAGCTGTATTTCTTTATAACAACAATCAGAGCATATCGGGATGAC	503
Qy	140	IleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrrTrpSerCysCysGlyAlaArg	159
Db	504	ATTGAATTTGCAAAACCTCATAGACTTCACCCAGGAATATTTGGCAGTGCTGTGGGGCTTTT	563
Qy	160	GlyProAsnAspTrpAsnLeuAsnIleTyrrPheAsnCysThrAspLeuAsnProSerArg	179
Db	564	GGAGCTGATGATTGGAACTTAATATTTACTTCAATTCGACAGATTCCAATCGAAGTCGA	623
Qy	180	GluArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerTh	198
Db	624	GAGCGATGTGGCGTTCCATTCTCTGTCTGCACATAAAGATCCCGCAGAAGATGTCATCAAC	683
Qy	198	rProSerVallalaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrrH	218
Db	684	ACTCAGTGTGGCTATGATGCCAGGCAAAAAACAGAAAGTTGACCAGCAGATTGTAATCTAC	743
Qy	218	rProIysAlaValTrpAlaSerLeuArgSerGlyCysargThr	232
Db	744	ACGAAAGGCTGTGTGCCCGCCAGTTTTCAGAAGTGGTTGCAGGACA	786

RESULT 11

AA220854

ID AAZ20854 standard: cDNA; 1110 BP.

XX

AC AAZ20854;

XX  
XX

DT 09-DEC-1999 (first entry)

XX  
DE Polynucleotide sequence of the dk329 1 clone.

XX secreted protein; cDNA library; clone; transmembrane protein;  
KW signal sequence cloning; hybridization cloning; gene therapy;  
KW receptor; ds.

XX

OS Homo sapiens.

XX	Key	Location/Qualifiers
FH		

FT	CDS	176..757
----	-----	----------

```
FT / *tag= a
FT /product= dk329 1
```

FT	sig peptide	326..424
FT	sig peptide	326..424

```
FT
/*tag= b
```

FT mat\_peptide 425..757

FT /\*tag= c

XX

PN WO9942470-A1.

XX

PD 26-AUG-1999.

	•	•
	<	<
	•	•
	•	•
	{	{
	{	{
	[	[
	•	•
	•	•
⊗	⊗	⊗
⊗	⊗	⊗

PF 666T-83J-8T  
yy 18-FEB-1999;

XX  
DB 10 FEB 1968

PR 18-FEB-1998; 98US=00/5038.  
PR 17-FEB-1998; 98US=0351600

PR 17--FEB-1999; 99US-0251600.  
XX

PA (GEMV ) GENETICS INST INC

(GEM) / GENETICS INST INC.  
XX

PI Jacobs K. McCov JM. LaVall

PI Treacy M. Agostino MJ. S

XX  
XX

DR WPI; 1999-518580/43.

DR P-PSDB; AAY42381.

XX

**PT New polynucleotides**

PT therapeutic, diagnosis

—

Qy 181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199  
 Db 646 CGATGTGGCGTTCATTCTCTGCTGCACTAAAGATCCGACAGAGATGTCATCAACT 705  
 Qy 199 oSerValAlaMetMetSerGlySerAnThrSerTrpSerSerArgAlaProTyThrPr 219  
 Db 706 CAGTGTGGCTATGATGCCAGGCAAAACCAAGAGTTGACAGCAGATTTGTAATCTACAG 765  
 Qy 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
 Db 766 AAAGGCTGTGTGCCCGAGTTTGAGAGTGTGTCAGGACA 805  
 RESULT 12  
 AAS59291  
 ID AAS59291 standard; cDNA; 1110 BP.  
 XX  
 AC AAS59291;  
 XX  
 DT 16-JAN-2002 (first entry)  
 XX  
 DE Human cDNA encoding a secreted protein dk329\_1.  
 XX  
 KW Human; secreted protein; ss; antiinflammatory; immunosuppressive;  
 KW neurotropic; neuroprotective; antiarthritic; antimicrobial; vulnary;  
 KW cytotatic; antidiabetic; virucide; antifertility; anticonvulsant;  
 KW vasotropic; antiparkinsonian; immunostimulant; dermatological;  
 KW antirheumatic; antitumor; anticulcer; osteopathic; tranquiliser;  
 KW cerebroprotective; cytokine; cell proliferation; cell differentiation;  
 KW immune deficiency; severe combined immunodeficiency; SCID; tumour;  
 KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
 KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
 KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
 KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
 KW food supplement.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175068-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 22-MAR-2001; 2001WO-US09369.  
 XX  
 PR 30-MAR-2000; 2000US-0539330.  
 PR 04-DEC-2000; 2000US-0729674.  
 XX  
 PR (GEMY ) GENETICS INST INC.  
 XX  
 PA Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;  
 PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
 PI Clark H, Rechtel K, Merberg D;  
 XX  
 DR WPI; 2001-639363/73.  
 DR F-PSDB; AAU39073.  
 XX  
 PT Secreted human proteins, useful as vaccine for treating various  
 PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
 PT nervous system disorders (e.g. stroke) -  
 XX  
 PS Disclosure; Page 576-577; 619pp; English.  
 XX  
 CC The invention relates to novel human secreted proteins, the nucleic  
 CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
 CC or cell differentiation activity or may induce production of other  
 CC cytokines in certain cell populations and may exhibit immune stimulating  
 CC or immune suppressing activity, which is useful for the treatment of  
 CC various immune deficiencies and disorders e.g. severe combined  
 CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
 CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
 CC inflammation. The proteins are also useful in the treatment of diseases  
 CC and disorders including tissue, skin and organ transplantation and in  
 CC graft-versus-host diseases (GVHD), in the induction of tumour immunity.

CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
 CC in the treatment of burns, incisions and ulcers; as well as in treatment  
 CC of periodontal diseases, osteoporosis or osteoarthritis, mediated by  
 CC inflammatory processes, diseases of the peripheral nervous system,  
 CC Alzheimer's, Parkinson's disease, Huntington's disease,  
 CC ankytrophic lateral sclerosis, and Shy-Drager syndrome, infections,  
 CC infarction of cardiac and central nervous system vessel e.g. stroke,  
 CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
 CC protein, having activin- or inhibin-related activities is useful as a  
 CC contraceptive based on the ability of inhibins to decrease fertility in  
 CC female mammals and decrease spermatogenesis in male mammals. The  
 CC proteins and nucleic acids are also useful as food supplements. The  
 CC present sequence encodes a secreted protein of the invention.  
 XX  
 SQ Sequence 1110 BP; 261 A; 273 C; 288 G; 285 T; 3 other;

Alignment Scores:  
 Pred. No.: 8,61e-69 Length: 1110  
 Score: 807.00 Matches: 163  
 Percent Similarity: 76.50% Conservative: 16  
 Best Local Similarity: 69.66% Mismatches: 30  
 Query Match: 62.90% Indels: 26  
 DB: 22 Gaps: 3

US-09-972-970-4 (1-233) x AAS59291 (1-1110)

Qy 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20  
 Db 176 ATGTCCGGGAAG-----CACTACAAAGGGTCTGGAAGTCAGTTGTCATCAAAATCTTC 229  
 Qy 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 230 ATATTTGGCTTCAATGTCATATTTGGTTTGGGAATAACATTTCTTGGAAITGGACTG 289  
 Qy 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 290 TGGGCATGGAATGAAAGAGGAGTTCTGTCAACATCTTCCATCCAGTCGCGGCGC 349  
 Qy 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
 Db 350 TTGACCCAGTTTGGCTTCTTCTGTGGGGAGGAGTGATGTTTCATTTGGGATTTGCA 409  
 Qy 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
 Db 410 GGGTGCAATTGGAGCGCTACGGGAAACACTTTCCTTCTCAAGTTTTTCTGTGTTCCCTG 469  
 Qy 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
 Db 470 GGAATTTATTTTCTTCTCGAGCTCACTGCGGGAGTTCTAGCATTTGTTTCAAGACTGG 529  
 Qy 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle 140  
 Db 530 ATCAAAGACCAGCTGTATTTCTTTATAAACAACATCAGAGCATATCGGATGACATT 589  
 Qy 141 AspLeuGlnAsnLeuIleAspPheAlaGlnTyrTrpSerCysGlyAlaArgGly 160  
 Db 590 GATTTGCAAAACCTCATGACTTCAACCAGGAATATAT-TCC----- 630  
 Qy 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
 Db 631 -----AATGCAAGTCGAGAG 645  
 Qy 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199  
 Db 646 CGATGTGGCGTTCATTCTCTGCTGCACTAAAGATCCGACAGAGATGTCATCAACT 705  
 Qy 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyThrPr 219  
 Db 706 CAGTGTGGCTATGATGCCAGGCAAAACCAAGAGTTGACAGCAGATTTGTAATCTACAG 765  
 Qy 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
 Db 766 AAAGGCTGTGTGCCCGAGTTTGAGAGTGTGTCAGGACA 805







Db 646 CGATGTGGCGTTCCTCTGCTGCTAAAGATCCCGCAGAGATGTCATCAACACT 705  
QY 199 oSerValAlaMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219  
Db 706 CAGTGTGGCTATGATCCAGGCAAAACCAAGAGTTGACACAGAGATTGTAATCTACAG 765  
QY 219 olysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
Db 766 AAGGCTGTGTGCCCCAGTTTGAGAGTGGTTGCAGGACA 805

RESULT 14  
AAI72287  
ID AAI72287 standard; cDNA; 864 BP.  
XX  
AC AAI72287;  
XX  
DT 15-APR-2002 (first entry)  
XX  
DE NET-4 antisense molecule.  
XX  
KW NET-4; antisense; modulator; neoplastic disease; cell proliferation;  
KW gene expression; triple-helix; polymerase; transcription factor;  
KW cancer therapy; hyperproliferation; tumour; growth; invasion;  
KW metastasis; ss.  
XX  
OS Homo sapiens.  
XX  
WO200198350-A2.  
XX  
PD 27-DEC-2001.  
XX  
PF 05-JUN-2001; 2001WO-US18415.  
XX  
PR 07-JUN-2000; 2000US-209865P.  
XX  
PA (CHIR ) CHIRON CORP.

XX  
PI Reinhard C, Jefferson AB, Winter JA, Randazzo F;  
XX  
DR WPI; 2002-147795/19.  
XX  
XX  
XX Novel NET-4 modulator useful for decreasing expression of NET-4 in a  
PT mammalian cell and treating neoplastic disease, is selected from  
PT antisense oligonucleotide, ribozyme, protein, polypeptide and a small  
PT molecule -  
XX  
XX Claim 3; Page 50; 57pp; English.

XX This sequence represents a NET-4 antisense molecule which acts as a  
CC NET-4 modulator. The modulator is useful for decreasing the expression  
CC of NET-4 in a mammalian cell, and for treating neoplastic disease,  
CC such that the neoplastic disease is reduced in severity. Modulators  
CC of NET-4 are also useful for regulating cell proliferation, and for  
CC controlling gene expression through triple-helix formation which  
CC promotes the ability of the double helix to open sufficiently for  
CC the binding of polymerase, transcription factors or regulatory  
CC molecules. A NET-4 modulator is useful as drug for supplementing  
CC cancer therapeutics and other agents. It is also useful in  
CC other diseases of hyperproliferation, and to inhibit tumour cell  
CC growth, invasion or metastasis.

XX  
SQ Sequence 864 BP; 162 A; 229 C; 249 G; 224 T; 0 other;

Alignment Scores:  
Pred. No.: 2,768-67 Length: 864  
Score: 790.00 Matches: 139  
Percent Similarity: 88.51% Conservative: 15  
Best Local Similarity: 79.89% Mismatches: 18  
Query Match: 61.57% Indels: 2  
DB: 24 Gaps: 1

US-09-972-970-4 (1-233) x AAI72287 (1-864)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20  
Db 348 ATGTCCGGGAAG-----CACTACAAAGGGTCTGGAAGTCAGTTGGTCATCAAAATCTTC 401  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 402 ATATTGGCTTCAATGTCTATTTTGGTTTGGGAATAACATTTCTTGGAAATGGACTG 461  
QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 462 TGGGCATGGAATGAAAAGAGAGTCTGTCCAAACATCTCTTCCATCACCAGATCTCGCGGC 521  
QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 522 TTTGACCCAGTTGGCTCTTCTTGTGGTGGAGGAGTGTTCATTTTGGGATTGCA 581  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPheSerValPheLeu 100  
Db 582 GGGTGCATTGGAGCGCTACGGGAAAACACTTTCTTCTCAAGTTTCTGTGTCTCTG 641  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTyr 120  
Db 642 GGAATTATTCTTCTCTGGAGCTCACTGCCGAGTTCTAGCATTTCTTTTCAAAGACTGG 701  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValValSalatYrArgAspAlle 140  
Db 702 ATCAAAGACCAGCTGTATTCTTTTATAACAACAACATCAGAGCATATCGGGATGACAT 761  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaAArgGly 160  
Db 762 GATTTCGAAAACCTCATAGACTTCACCAGAGTAATTTGCGAGTGTCTGTGGGCTTTTGA 821  
QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp 174  
Db 822 GCTGATGATTGGAACTAAATATTACTTCAATTGCACAGAT 863  
RESULT 15  
ABK35735  
ID ABK35735 standard; cDNA; 1988 BP.  
XX  
AC ABK35735;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE cDNA sequence #126 encoding novel human secreted protein.  
XX  
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
KW immune deficiency disorder; blood disorder; inflammatory disorder;  
KW infectious disorder; allergic condition; neurodegenerative disorder;  
KW liver fibrosis; coagulation disorder; gene therapy; antimicrobial;  
KW tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.  
OS Homo sapiens.  
XX  
WO200177289-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 29-MAR-2001; 2001WO-US10232.  
XX  
PR 06-APR-2000; 2000US-195605P.  
XX  
PA (GEMY ) GENETICS INST INC.  
XX  
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;  
PI Clark HF, Fehnel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;  
DR WPI; 2002-179322/23.  
XX  
XX Six hundred and twenty three polynucleotides derived from a variety of  
PT human tissue sources which encode secreted proteins, useful for  
PT treating immune deficiencies and disorders such as autoimmune disorders

PT

XX Claim 1; Page 146; 393pp; English.

Search completed: November 21, 2003, 15:34:34  
Job time : 285 secs

CC The present invention relates to the isolation of novel cDNA sequences  
CC which encode human secreted proteins. The cDNA sequences have been  
CC derived from a variety of human tissues. The invention also provides  
CC a method for producing proteins from these polynucleotide sequences.  
CC The proteins are useful for identifying compounds that modulate their  
CC activity and production. The sequences of the invention are  
CC useful for treating diseases such as hyperproliferative disorders  
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined  
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple  
CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory  
CC disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),  
CC allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.  
CC Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.  
CC haemophilia), and tumours. The polynucleotide sequences of the  
CC invention are also useful in gene therapy. ABK35610-ABK36232 represent  
CC the cDNA sequences of the invention that encode for novel human  
CC secreted proteins.

XX  
SQ Sequence 1988 BP; 503 A; 434 C; 446 G; 604 T; 1 other;

## Alignment Scores:

Pred. No.:	7,65e-64	Length:	1988
Score:	759.50	Matches:	146
Percent Similarity:	83.33%	Conservative:	14
Best Local Similarity:	76.04%	Mismatches:	30
Query Match:	59.20%	Indels:	2
DB:	24	Gaps:	1

US-09-972-970-4 (1-233) x ABK35735 (1-1988)

QY	43	TrpGlyGluIysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGlyLeuAsp	62
DB	2	TGGAAATGAAAGAGTTCTGTCCAAACATCTTCATCACCAGATCTCGCGGCTTTGAC	61
QY	63	ProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAlaGlyCys	82
DB	62	CCAGTTTGGCTCTTCCTGTGTGGGAGGATGATGTTTCATTTGGGATTTGCAGGGTGC	121
QY	83	IleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeuGlyLeu	102
DB	122	ATTGGAGCGCTACGGGAAACACTTTCCTTCCTCAAGTTTTTTCGTGTTCTCGGAATT	181
QY	103	IlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrpIleArg	122
DB	182	ATTTTCTCTCGAGCTCACTCGCGGAGTTCTAGCATTTGTTTCAAAGACTCGATCAAA	241
QY	123	AspGlnLeuAsnPhePheIleAsnAsnValIysAlaTyrArgAspAspIleAspLeu	142
DB	242	GACCACTGTATTTCTTTTATAACAACAACACAGACATATCGGGATGACATGATTTG	301
QY	143	GlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGlyProAsn	162
DB	302	CAAAACCTCATAGACTTACCCAGGAATATTGGCAGTGTGTGGGGCTTTTCGAGCTGAT	361
QY	163	AspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGluArgCys	182
DB	362	GATTGGAACCTAAATATTACTTCAATTGTCGCCAGAGATGTATCAACACTCAGTGT	421
QY	183	GlyValProPheSerCysCysValArgAspProAla---MetSerSerThrProSerVa	201
DB	422	GGGGTCCATTCTCTGTGTCACATAAGATCCCGAGAGATGTATCAACACTCAGTGT	481
QY	201	IleMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyrThrProLysAl	221
DB	482	GGCTATGATGCCAGGCAAAACAGAAAGTTGACACAGATTGTAATCTACACGAAGGC	541
QY	221	aValTrpAlaSerLeuArgSerGlyCysArgThr	232
DB	542	TGTGTGCCCCCGTTTGAGAAGTGGTTGCAGGACA	575

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2003, 13:58:33 ; Search time 3729 Seconds  
(without alignments)  
2556.166 Million cell updates/sec

Title: US-09-972-970-4

Perfect score: 1283

Sequence: 1 MPQKHQHFQEPVEGCGKYF.....RAPYTPKAVWASLRSGCRTT 233

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09972970/runat\_21112003\_125335\_27229/app\_query.fasta\_1.391  
-DB=genEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORES=ext -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-USER=US09972970 @CGN 1 1 3508 @runat 21112003 125335 27229 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

RESULT 1

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.nam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1215.5	94.7	813	6	AX420468	AX420468 Sequence
2	1215.5	94.7	2426	9	BC010405	BC010405 Homo sapi
3	1215.5	94.7	3184	6	AX420466	AX420466 Sequence
4	1174.5	91.5	1516	10	BC010346	BC010346 Mus muscu
5	958.5	74.7	1405	9	AF065389	AF065389 Homo sapi
6	958.5	74.7	1416	9	BC009704	BC009704 Homo sapi
7	955.5	74.5	1567	6	AX061778	AX061778 Sequence
8	954.5	74.4	1685	5	BC041304	BC041304 Xenopus l
9	944.5	73.6	3175	10	AF121344	AF121344 Mus muscu
10	925.5	72.1	1408	9	AF053455	AF053455 Homo sapi
11	853.5	66.5	4445	9	AK024427	AK024427 Homo sapi
12	807	62.9	1110	6	BD135990	BD135990 Secretary
13	790	61.6	864	6	AX343015	AX343015 Sequence
14	671	52.3	2428	10	BC025568	BC025568 Mus muscu
15	671	52.3	2498	10	BC024611	BC024611 Mus muscu
16	671	52.3	2500	10	BC026574	BC026574 Mus muscu
17	654.5	51.0	1388	6	AX440923	AX440923 Sequence
18	654.5	51.0	2553	9	AF311903	AF311903 Homo sapi
19	651.5	50.8	2556	9	HS080161	AL136638 Homo sapi
20	579	45.1	1998	3	AK116798	AK116798 Ciona int
21	570	44.4	2502	9	BC002920	BC002920 Homo sapi
22	537	41.9	368	9	AF174603	AF174603 Homo sapi
23	501	39.0	171419	2	AC135142	AC135142 Rattus no
24	501	39.0	234117	2	AC130985	AC130985 Rattus no
25	489	38.1	187478	2	AC123758	AC123758 Mus muscu
26	489	38.1	215066	2	AC136719	AC136719 Mus muscu
27	476.5	37.1	852	6	AX247836	AX247836 Sequence
28	476.5	37.1	2073	9	BC044244	BC044244 Homo sapi
29	474.5	37.0	1995	10	BC024685	BC024685 Mus muscu
30	440.5	34.3	132832	9	AC091934	AC091934 Homo sapi
31	430	33.5	140468	2	AC017377	AC017377 Drosophil
32	430	33.5	174311	3	AC004758	AC004758 Drosophil
33	430	33.5	174832	3	AC092216	AC092216 Drosophil
34	430	33.5	260673	3	AE003612	AE003612 Drosophil
35	426.5	33.2	168318	2	AC024042	AC024042 Homo sapi
36	421.5	32.9	1816	3	BT004914	BT004914 Drosophil
37	412.5	32.2	1216	3	AF274013	AF274013 Drosophil
38	404.5	31.5	22920	3	CET14G10	Z68880 Caenorhabdi
39	402	31.3	102468	2	AC009984	AC009984 Drosophil
40	402	31.3	106584	2	AC013977	AC013977 Drosophil
41	402	31.3	140973	3	AC008140	AC008140 Drosophil
42	402	31.3	159007	3	AC009219	AC009219 Drosophil
43	402	31.3	225594	3	AE003688	AE003688 Drosophil
44	373.5	29.1	102298	10	AL671671	AL671671 Mouse DNA
45	367	28.6	267	6	BD072489	BD072489 Secreted

ALIGNMENTS

AX420468	AX420468	813 bp	DNA	linear	PAT 18-JUN-2002
LOCUS	Sequence 3 from Patent WO0216603.				
DEFINITION	AX420468				
ACCESSION	AX420468				
VERSION	AX420468.1	GI:21524616			
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1				
AUTHORS	Leiby, K.R.				
TITLE	23228, a human tetraspanin family member and uses thereof				
JOURNAL	Patent: WO 0216603-A 3 28-FEB-2002;				
JOURNAL	Millennium Pharmaceuticals, Inc. (US)				
FEATURES	Location/Qualifiers				
source	1..813				
BASE COUNT	151 a	233 c	235 g	194 t	
ORIGIN	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
Alignment Scores:	9.98e-96	Length:	813		
Pred. No.:	1215.50	Matches:	227		
Score:	96.60%	Conservative:	0		
Percent Similarity:	96.60%	Mismatches:	6		
Best Local Similarity:	96.60%	Indels:	2		
Query Match:	94.74%	Gaps:	1		
DB:	6				
US-09-972-970-4 (1-233) x AX420468 (1-813)					
QY	1	MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe	20		
DB	1	ATGCCCGGCAAGCAGCAGCATTTTCAGGAACCTGAGTGGCTGGTGGGGAAATACATTC	60		
QY	21	LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu	40		
DB	61	CTGTTTGGCTTCAACATGCTCTCTGGTGCTGGAGCCCTGCTCTGCTATCGCCCTC	120		
QY	41	TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGly	60		
DB	121	TGGGGCTGGGGTGAAGAGCGCTTCTCGAATCATCTCAGCGCTGACATCTGGAGGC	180		
QY	61	LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla	80		
DB	181	CTTGACCCCGTGGCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	240		
QY	81	GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu	100		
DB	241	GGCTGCATTGGGGCCCTCCGGGAGAACACCTTCTCTGCTCAAGTTTCTCCGTGTCCTC	300		
QY	101	GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp	120		
DB	301	GGTCTCATCTTCTCTGGAGTGGCAACAGGATCTTGGCCCTTGTCTTCAAGGACTGG	360		
QY	121	IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle	140		
DB	361	ATTGAGACCACTCACTCTTTCATCAACAACACGTCAGGCTTACCGGACGACATT	420		
QY	141	AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly	160		
DB	421	GACCTCCAGAACCTCATTTGACTTCTCTCAGGAATACTGTTCTGTGTGGAGCCGAGGC	480		
QY	161	ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu	180		
DB	481	CCCAATGATCGAACCTCAATATCTTACTTCACTGCACTGACTTGAACCCAGCGGAG	540		
QY	181	ArgCysGlyValProPheSerCysCysValArgAspProAla---MetSerSerThrPr	199		
DB	541	CGCTGGGGGGTCCCTTCTCTGCTGCTGAGGACCTCGGAGGATGTCTCTCAACACC	600		

Qy	199	oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr	219		
Db	601	CAGTGTGGCTTACGACGCTCGGCTCAAACTGGAGCTGGAGCAGCAGGAGTTCATCCACACC	660		
Qy	219	olybAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr	233		
Db	661	AAAGGCTGGCTGGCCAGTTTGAGAAAGTGGCTGCAGGACACACC	703		
RESULT 2					
BC010405	2426 bp	mRNA	linear	PRI 10-JUN-2003	
LOCUS	Homo sapiens hypothetical protein MGC14859, mRNA (cDNA clone				
DEFINITION	MGC:14859 IMAGE:3621871), complete cds.				
ACCESSION	BC010405				
VERSION	BC010405.1	GI:14714540			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1	(bases 1 to 2426)			
AUTHORS	Strausberg, R.L., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udutin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalobon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzyzanski, M.I., Skalska, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.				
TITLE	human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
MEDLINE	22388257				
PUBMED	12477932				
REFERENCE	2	(bases 1 to 2426)			
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>				
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology <a href="http://www.systemsbio.org">http://www.systemsbio.org</a> contact: <a href="mailto:amadan@systemsbiology.org">amadan@systemsbiology.org</a> Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting				
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>					
Series: IRAL Plate: 24 Row: k Column: 15					
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.					
FEATURES	Location/Qualifiers				
source	1..2426				
	/organism="Homo sapiens"				



QY	101	GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheIleAspTrp	120
Db	468	GGTCTCATCTTCTTCTGAGCTGGCAACAGGATCTGGCCCTTCTTCAAGGACTGG	527
QY	121	IleArgAspGlnLeuAsnPhePheIleAsnAsnValIleAlaTyrArgAspAspIle	140
Db	528	ATTGAGAGCAGCTCAACCTTCTCATCAACACAGCTCAAGCCCTACCGGACGACATT	587
QY	141	AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly	160
Db	588	GACCTCCAGAACCTCATTTGCTTGTCTGAGGATCTGCTTCTGCTGTGGAGCCGAGGC	647
QY	161	ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu	180
Db	648	CCCAATGACTGGAACTCAATATCTACTTCAACTGCACCTGACTTGAACCCCGAGGGAG	707
QY	181	ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr	199
Db	708	CGCTGGGGGTGCCCTTCTCTGCTGGTCAAGGACCTCGGAGGATGCTCTCAACACC	767
QY	199	oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr	219
Db	768	CAGTGTGGCTACGACTCGCGCTCAAACTGGAGCTGGAGCAGCAGGCTTCATCCACACC	827
QY	219	oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr	233
Db	828	AAAGGCTGGTGGCCAGTTTGAAGATGGCTGCAGGACAACC	870
RESULT 4			
BC010346			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			

101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheIleAspTrp 120

468 GGTCTCATCTTCTTCTGAGCTGGCAACAGGATCTGGCCCTTCTTCAAGGACTGG 527

121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValIleAlaTyrArgAspAspIle 140

528 ATTGAGAGCAGCTCAACCTTCTCATCAACACAGCTCAAGCCCTACCGGACGACATT 587

141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly 160

588 GACCTCCAGAACCTCATTTGCTTGTCTGAGGATCTGCTTCTGCTGTGGAGCCGAGGC 647

161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180

648 CCCAATGACTGGAACTCAATATCTACTTCAACTGCACCTGACTTGAACCCCGAGGGAG 707

181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199

708 CGCTGGGGGTGCCCTTCTCTGCTGGTCAAGGACCTCGGAGGATGCTCTCAACACC 767

199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219

768 CAGTGTGGCTACGACTCGCGCTCAAACTGGAGCTGGAGCAGCAGGCTTCATCCACACC 827

219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233

828 AAAGGCTGGTGGCCAGTTTGAAGATGGCTGCAGGACAACC 870

BC010346 1516 bp mRNA linear ROD 16-APR-2003

Mus musculus RIKEN cDNA 2210021G21 gene, mRNA (cDNA clone MGC:6941 IMAGE:2811935), complete cds.

BC010346

MGC.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1516)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E., Sutterich, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 1516)

Strausberg, R.

Direct Submission

Submitted (05-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [angbcm.tmc.edu](mailto:angbcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK plate: 5 Row: f Column: 18.

Location/Qualifiers

1..1516

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="MGC:6941 IMAGE:2811935"

/tissue\_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."

/clone\_lib="NCI CGAP\_Mam5"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..1516

/gene="2210021G21Bik"

/db\_xref="locusID:74257"

/db\_xref="MGI:1921507"

136..948

/codon\_start=1

/product="RIKEN cDNA 2210021G21"

/protein\_id="AAH10346.1"

/db\_xref="GI:16307593"

/db\_xref="LocusID:74257"

/translation="MPGKHQFQDPEVGVCCGKYFLGFNVPWVILGALFLAIGLWANG EKVLSNLSALTDLGGDPVWLVFVVGWMSVLGFGAGCIGALRENTFLKPFVSFLGL IFFLEAAGILAFVKDRIQDLNFIINNKAAYRDDLDLQNLIDPAQYWSCCGARG PNDNLNLYFNCTDLNRSRRCGVPFSCVDPADVLNQTGQYDIRLKLLEQQQSGI YTKGCVGFQKMLQNLIVAGLVGIALQLIFGLCLAQNLVSDIKAVKANW"

BASE COUNT 277 a 417 c 461 g 361 t

ORIGIN

Alignment Scores:

Pred. No.: 7,12e-92 Length: 1516

Score: 1174.50 Matches: 221

Percent Similarity: 95.32% Conservative: 3

Best Local Similarity: 94.04% Mismatches: 9

Query Match: 91.54% Indels: 2

DB: 10 Gaps: 1

US-09-972-970-4 (1-233) x BC010346 (1-1516)

QY 1 MetProGlyIleHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20

Db 136 ATGCCCGCAAGCACCAGCAATTCAGGACCTCGAGTCTGCTGCGGGAATCTTC 195

QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40

Db 196 CTGTTTGGCTTCAACATTGCTCTCTGGTGTGGAGCCCTGTTCTCGGCCATTGGCCTC 255

QY 41 TrpAlaTrpGlyGluIleValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60

Db 256 TGGGCTGGGGTGAAGAGGGGCTTCTCTCAACATCTCGGCGCTGACAGATCTGGCGCGT 315

QY 61 LeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGlyPheAla 80

Db 316 CTTGACCCCGTGGCTGTTTGTGGTGGTGGGGGAGTCAATGTCAGTGTGGCTTGGCCTTCC 375

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPheSerValPheLeu 100  
Db 376 GGCCTGATGGGGCCCTCGGGAAACACCTTCCTGCTCAAGTTCTTCTGTGTCTC 435  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 436 GGCCTCATCTTCTTCTGAGCTGGCCGGGGATCCTGGCTTCTTCAAGATTGG 495  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyArgAspSepile 140  
Db 496 ATCCGAGACCAACTTAATCTTCTCAACAACAATGTCAAAGCCTACCGGATGATCTC 555  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysGlyAlaArgGly 160  
Db 556 GACCTTCAGAACCTTATCGACTTCTCAGGAATACTGCTCTTGTGGTCCCGAGGG 615  
QY 161 ProAsnAspTrpAsnLeuAsnIleTyPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 616 CCCAATGATCGAACCTCAATATCTACTTCACTGACTGACTGACTGAAACCCAGCGTGAG 675  
QY 181 ArgCysGlyValProPheSerCysValArgAspProAla---MetSerSerThrPr 199  
Db 676 CGCTGTGGGTGCCCCCTTCTGCTGTGTAGGGACCCAGCGGAAGATGTCCTCAATACC 735  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyThrPr 219  
Db 736 CAGTGTGGCTATGACATCCGACTCAAACTGGAGCTCGAGCAGCAGGGCTCCATCACC 795  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233  
Db 796 AAAGGCTGTGTGGGCCAGTTTGGAGAAGTGGCTACAAGACAACC 838

RESULT 5  
AF065389 1405 bp mRNA linear PRI 28-APR-2000  
LOCUS Homo sapiens tetraspan NET-4 mRNA, complete cds.  
DEFINITION AF065389  
VERSION AF065389.1 GI:3152702  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1405)  
AUTHORS Serru.V., Dessen.P., Boucheix.C. and Rubinstein.E.  
TITLE Sequence and expression of seven new tetraspans  
JOURNAL Biochim. Biophys. Acta 1478 (1), 159-163 (2000)  
MEDLINE 20185353  
PubMed 10719184  
REFERENCE 2 (bases 1 to 1405)  
AUTHORS Rubinstein.E., Serru.V. and Boucheix.C.  
TITLE Direct Submission  
JOURNAL Submitted (14-MAY-1998) INSERM U268, 14 av Paul Vaillant Couturier,  
Villejuif 94807, France  
FEATURES  
source Location/Qualifiers  
1. .1405  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE ID 219547"  
348. .1154  
/note="similar to Tspan-5; TM4SF"  
/codon\_start=1  
/product="tetraspan NET-4"  
/protein\_id="AAC17120.1"  
/db\_xref="GI:3152703"  
/translation="MSGKHVGPESVCCIKYFIFGFNFIWFLGILFLGLWAWNEK  
GVLSNTSSITDGLGDFPVMFLVVGVMFIFLPGAGCIGALRENTLLKPFVSFLGIIF  
FLELTAGVLAFLVKDWIKDLYFNNIYAVRDDIDLDNLIDFTQIEYWCQCAFAGD  
DWNLIYFNCTDSNASRRCGVPSCTKDPADVDINTQCGYDARQKPEVDQIVIT  
KGCVPQEFKWLQNLITVAGIFIGIALIQLFICLQNLVSDIEAVRASW"

BASE COUNT 309 a 357 c 387 g 352 t  
ORIGIN

Alignment Scores: 2.6e-73 Length: 1405  
Pred. No.: 958.50 Matches: 180  
Score: 84.19% Conservative: 17  
Best Local Similarity: 76.92% Mismatches: 33  
Query Match: 74.71% Indels: 4  
DB: 9 Gaps: 2  
US-09-972-970-4 (1-233) x AF065389 (1-1405)  
QY 1 MetProGlyLysHisGlnPheGlnGluProGluValGlyCysCysGlyLysTyPhe 20  
Db 348 ATGTCGGGAAG-----CACTACAGGGTCTCTGAAGTCAGTTGTTCATCAATACTTC 401  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 402 ATATTGGCTCAATGTATATTTGGTTTGGGAATAACATTCTTGGATTGGACGTG 461  
QY 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 462 TGGGATGGAATGAAAGAGGAGTTCTGTCCAACATCTCTTCATCACCAGATCTCGCGGC 521  
QY 61 LeuAspProValTrpLeuPheValValGlyValMetSerValLeuGlyPheAla 80  
Db 522 TTTGACCCAGTTTGCTCTTCTGTGGTGGAGGATGATGTTTCATTTTGGGATTGCA 581  
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
Db 582 GGGTGATTTGGAGCGCTACGGGAAACACTTCTCTCAAGTTTCTTCTGTGTCTCTG 641  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 642 GGAATTTATTTCTTCTGGAGTCTCACTGCCGGAGTTCTAGCATTTGTTTCAAGACTGG 701  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyArgAspSepile 140  
Db 702 ATCAAGACCAGCTGTATTTCTTTATTAACAACATCAGAGCATATCGGGATGACATT 761  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysGlyAlaArgGly 160  
Db 762 GATTTCAAAACCTCATAGACTTCACCAGGAATATTGGCAGTCTGTGGGGCTTTTGA 821  
QY 161 ProAsnAspTrpAsnLeuAsnIleTyPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 822 GCTGATGATTGGAAACCTAAATATTACTTCAATTGACAGATTCCAATGCAAGTCGAGAG 881  
QY 181 ArgCysGlyValProPheSerCysValArgAspProAla---MetSerSerThrPr 199  
Db 882 CGATGTGGCGTTCATTTCTCTGCTGCACTAAAGATCCCGCAGAGATGTTCATCAACT 941  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerArgAlaProTyThrPr 219  
Db 942 CAGTGTGGCTATGATGCCAGGCAAAACAGAGTTGACAGCAGATGTTAATCTACAG 1001  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
Db 1002 AAAGGCTGTGTGGCCCCAGTTTGAAGAGTGGTTGAGGACA 1041  
RESULT 6  
LOCUS BC009704 1416 bp mRNA linear PRI 13-JAN-2003  
DEFINITION Homo sapiens, tetraspan 5, clone MGC:9300 IMAGE:3895933, mRNA,  
complete cds.  
ACCESSION BC009704  
VERSION BC009704.1 GI:16307230  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1416)  
AUTHORS Strausberg,R.

TITLE  
JOURNAL

Direct Submission  
Submitted (29-JUN-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-3590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: AICC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arranged by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdpaxil.stanford.edu](mailto:mcdpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 14 Row: h Column: 4  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 21264582.  
Location/Qualifiers  
1..1416  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:9300 IMAGE:3895933"  
/tissue\_type="Pancreas, epithelioid carcinoma"  
/clone\_lib="NIH MGC\_70"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
334..1140  
/codon\_start=1  
/product="tetraspan 5"  
/protein\_id="AAH09704.1"  
/db\_xref="GI:16307231"  
/db\_xref="LocusID:10098"  
/translation="MSGKHVGPVEVSCIKYIFGFNFIWFLGTFIGIGLWANEK  
GVLSNITDGGDFVWLVGGVNFILGFAGCIGALRENTFLKFFSVFLGIIIF  
FLEITAGVLAFVKDWIKDQLYFFINIRAYRDDIDQLNIDFTQKYQCCGAFAD  
DWNLIYFNCTDSNARCRGCFSCCTKPAEDVINTCGVDARQKEVQQIVYIT  
KGCVPQEPKWLQDNLTVAGIFIGIFALLQIFGICLQNLVSDIEAVRASW"

FEATURES  
source

## CDS

BASE COUNT  
ORIGIN

## Alignment Scores:

Pred. No.: 2..62e-73 Length: 1416  
Score: 958.50 Matches: 180  
Percent Similarity: 84.1% Conservative: 17  
Best Local Similarity: 76.9% Mismatches: 33  
Query Match: 74.71% Indels: 4  
Dbs: 9 Gaps: 2

US-09-972-970-4 (1-233) x BC009704 (1-1416)

QY 1 MetProGlyLysHisGlnPheGlnGluProGluValGlyCysGlySerPhe 20  
Db 334 ATGTCGGGAG-----CACTACAAAGGCTCTGAAGTCAGTTGTGCATCAATACTTC 387  
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
Db 388 ATATTGGCTTCAATGTCATATTGGTTTGGGAATAACATTCTTGGAAATGGACTG 447  
QY 41 TrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
Db 448 TGGGATCGAATGAAAGAGAGTCTGTCCCAACATCTCTCCATCCCGATCTCGCGGC 507  
QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
Db 508 TTTGACCCAGTTTGGCTCTCTCTGTGGGAGGAGTGATGTTCAATTTGGGATTTGCA 567

QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuPhePheSerValPheLeu 100  
Db 568 GGGTCATTTGGAGCGCTACGGGAAACACTTTCTTCTCAAGTTTTTCTGTGTCTCTG 627  
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120  
Db 628 GGAATTATTTTCTTCTGAGCTCACTGCCGAGTCTTACGATTTGTTTCAAGACTGG 687  
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaIleArgAspIle 140  
Db 688 ATCAAAACACAGCTGATTTCTTTATAAACAACAACATCCAGAGCATATCGGATGACATT 747  
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpSerCysGlyAlaArgGly 160  
Db 748 GATTTCGAAAACCTCATAGACTTACCAGGAATATTCGAGTGTGTGGGCTTTTGA 807  
C/ 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
Db 808 GCTGATGATTGGAACCTAATATTACTTCAATTGCACAGATTCCTCAATGCAAGTCGAG 867  
QY 181 ArgCysGlyValProPheSerCysValArgAspProAla----MetSerSerThrPr 199  
Db 868 CGATGCGGTTCATCTCTGCTGCACTAAAGATCCGCAAGATGTCATCAACACT 927  
QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyThrPr 219  
Db 928 CAGTGTGCTATGATGCCAGGCAAAACCAAGTTGACCAAGCAGATTGTAAATCTACACG 987  
QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
Db 988 AAAGCTGTGTGCCCCCACTTTGAGAAAGTGGTTGCAGGACA 1027  
RESULT 7  
AX061778  
LOCUS AX061778 1567 bp DNA linear PAT 24-JAN-2001  
DEFINITION Sequence 2 from Patent WO0078948.  
ACCESSION AX061778  
VERSION AX061778.1 GI:12539860  
KEYWORDS Rattus sp.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1  
AUTHORS Andriamanandry, C. and Maitre, M.  
TITLE Cloning, expression and characterisation of a cDNA coding for a rat  
brain gamma-hydroxybutyrate (ghb) receptor  
JOURNAL Patent: WO 0078948-A 2 28-DEC-2000;  
Universite Louis Pasteur de Strasbourg (FR)  
FEATURES  
source Location/Qualifiers  
1..1567  
/organism="Rattus sp."  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10118"  
22..1560  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC25029.1"  
/translation="MRAAAGLPGRLHWAARRGLGRFPARVPRAAGRSPCPASIS  
NSRTRLAAGNTFCLASTLSSGCEPCSWPSAGPGVRRAPFFPTSGGQIQGGIDPV  
WLFVIGIMSVLGFAGCIGALRENTFLKFFSVFLGIIFFLELAAGTIAFVKDWIR  
DQLNLIINNKKAYRDDIDQLNIDFAQYSCCGAGPNDWNLIIRTSALTSPSR  
ERCGVFFCVVTRTKRSTQYPCGYTSQAELEHKLHLHOSWPKFKLQKDPGGLG  
AIAEFKGIAGPNRSTSRPRTGKLPGWATCLAPRAVVGASKDALPHSQWQGLWD  
VYDCLRLCEHTQATEAGLGNWCTAGLFTTELGVHMCVCVCVCVCR  
VCLCLVVRGMHVCAVSTPLVVRWPFOSFKGDGVRVHVPASVLLWDV  
SLGCLGACCLRLPHIHDLHPAWSPWQCHSLEMGRILSVLSRLPLRTHWQDGY  
ERLN"  
BASE COUNT 281 a 436 c 480 g 370 t  
ORIGIN



Alignment Scores:

Pred. No.:	5,36e-73	Length:	1567
Score:	955.50	Matches:	193
Percent Similarity:	85.11%	Conservative:	29
Best Local Similarity:	82.13%	Mismatches:	7
Query Match:	74.47%	Indels:	8
DB:	6	Gaps:	1

US-09-972-970-4 (1-233) x AX061778 (1-1567)

QY 1 MetProGlyHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 DB 135 ATGCCCGGAAGCATCAGCAATTCAGAGCCCTGAGCTGGCTGCTGCGGAAATACCTTT 194

QY 21 LeuPheGlyPheAsnIleValPheTyrValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 DB 195 CTGTTTGGCTTCAACATTGCTTCTGGGTGCTGGAGCCCTGTTCTGCGCATCGGCCCTC 254

QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAsp-LeuGlyGly 60  
 DB 255 TGGCCCTGGGTGAGAGGCGTCTTCTCAACATCTCAGGCGGACAGATCCAGGCGG 314

QY 60 YLeuAspProValTrpLeuPheValValGlyValGlyValMetSerValLeuGlyPheAl 80  
 DB 315 TCTTGACCCCGTGTGCTGTTTGTGTGATTTGGGGAATCATGTCAGTGTGCGCTTTGC 374

QY 80 aGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLe 100  
 DB 375 CGGTGTCATTTGGGCGCTCCGGGAAACCTTCTGCTCAAAATTTTCTGTGTCTCT 434

QY 100 uGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTr 120  
 DB 435 CGGCTCATCTTCTCTCGAGCTGGCGCGCGGATCTTGGCTTGTGTTCAAGGATTG 494

QY 120 PileArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspTr 140  
 DB 495 GATCCGAGACCACTTAACCTCTTCAATCAACAACATGTCAAAGCCCTACCGGACGAT 554

QY 140 eAspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyValaArgGly 160  
 DB 555 TGACCTTCAGACCTTATCGACTTCTGCTCAGAACTACTGCTTGTGTGGAGCCGAGG 614

QY 160 yProAsnAspTrAsnLeuAsnIleTyr--PheAsnCysThrAspLeu-AsnProSerAr 179  
 DB 615 GCCCAATGACTGGAACCTCAACATCGGACTTCACTGCACTGACTTCAAAACCCAGCG 674

QY 179 gGluArgCysGlyValProPheSerCysValArgAspProAlaMetSerSerThrPr 199  
 DB 675 CGAGCGCTGTGGGTGCCCTTCTGCTGGGTAAAG-GACCCTGGGGAAGAGCTCTCAAT 733

QY 199 oSerValAlaMetSerGlySerAsnTrpSerTrpSerArgAlaProTyrThrPr 219  
 DB 734 ACCCATGTGGCTATACATCCGCTCAAACTG-AGCTGGAGCAACAAGATTCATCTACACCA 792

QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArg 231  
 DB 793 AGCTGCT-----GGCCATTTCAGAGTGGCTCAAGA 823

RESULT 8  
 BC041304  
 LOCUS  
 DEFINITION  
 XENOPUS LAEVIS, Similar to transmembrane 4 superfamily member 9,  
 clone IMAGE:4683897, mRNA, partial cds.  
 ACCESSION  
 BC041304  
 VERSION  
 BC041304.1 GI:27735442  
 KEYWORDS  
 XENOPUS LAEVIS (African clawed frog)  
 XENOPUS LAEVIS  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus.  
 SOURCE  
 Klein, S. and Strausberg, R.

Direct Submission  
 Submitted (16-DEC-2002) National Institutes of Health, Xenopus Gene  
 Collection (XGC), National Institute of Child Health and Human  
 Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD  
 20892-7510, USA  
 NIH-MGC Project  
 Contact: XGC help desk  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. Igor Dawid  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
 http://www.systemsbio.org  
 contact: anadan@systemsbiology.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAC Plate: 94 Row: 9 Column: 18  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Similarity but not  
 identity to protein.

FEATURES  
 Location/Qualifiers  
 1..1685  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:4683897"  
 /tissue\_type="Embryo, stage 31/32, Xenopus"  
 /clone\_lib="NICHG\_XGC\_Emb4"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 <1..926  
 /codon\_start=3  
 /product="Similar to transmembrane 4 superfamily member 9"  
 /protein\_id="AAH41304.1"  
 /db\_xref="GI:27735443"  
 /translation="FGSTRVGLIYCFHLPRAVSPDPDPPRCQGNLPSCSGSMGKH  
 YKPEVSCCIYFIFFGNVIFMLGLTFLVGLWAMSEKVLNLSITDGLGFDPM  
 LFLVGVGVMIFLFGAGCTIAGTFLKFFSVFLGIIFLFLTAGVLAFFVKDWIKD  
 LQFFINNIRARDIDILQNLIDFTQVWCCGAGFADDMNLNLYFNCTDSNASRER  
 CGVPSCCTKPAEDVINTQGVYDVRKPELDQOETHTKGCVPQFEKWLQDNLTIVA  
 GVFTGIALQLIFGICLAQNLVSDIRAVRASW"

BASE COUNT 456 a 332 c 383 g 514 t

ORIGIN

Alignment Scores:

Pred. No.:	7,13e-73	Length:	1685
Score:	954.50	Matches:	175
Percent Similarity:	84.26%	Conservative:	23
Best Local Similarity:	74.47%	Mismatches:	33
Query Match:	74.40%	Indels:	4
DB:	5	Gaps:	2

US-09-972-970-4 (1-233) x BC041304 (1-1685)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe 20  
 DB 120 ATGTCGCGCAAG-----CACTACAAGGAGCCGAGGTCAGCTGTCATCAAGTACTTC 173

QY 21 LeuPheGlyPheAsnIleValPheTyrValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 DB 174 ATATTCGGCTTCAATGTCATCTTCTGGCTCTTGGACTTACGTTTCTTGGAGTTGCTGTG 233

QY 41 TrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 DB 234 TGGCATCGAGTGAAGAGGGTGTCTTCCAACTTCTCTCATCACAGATCTTGGGGGT 293

QY 61 LeuAspProValTrpLeuPheValValGlyValMetSerValLeuGlyPheAla 80  
 DB 294 TTTGATCCAGTGTGCTGTTTCTGTGTAGTCGAGGGGTGATGTTTATCTCTGGGCTTTGCA 353

```

QY 81 GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePheSerValPheLeu 100
Db 354 GCGTGTATTGGGGCATCTCGTGAGAACACCTTCTCTGTAATTTTTTCTGTGTTCTT 413

QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTTP 120
Db 414 GGAATTATTCTTCTGGAGCTCACGGGGAGTTCTTGCCTTCGTGTTCAAGACTGG 473

QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValIleAlaTyrArgAspAspIle 140
Db 474 ATTAAGACCACTGCGAGTCTTTATTAAACAACATCAGAGCCTACAGAGATGATC 533

QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160
Db 534 GACTTACAGAACCTTATTGCTTACAGAGGAATATTGCGAGTCTGCGGGGGTTCGGA 593

QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180
Db 594 GCTGATGATTGGAATTTAAACATTTATTTCACTGCACCGACTCTTAACGCGGAGAG 653

QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199
Db 654 AGATGTGGGTGCCATTTTCTGCTGCACAAAGACCTCTGAGGATGTAATAACACA 713

QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219
Db 714 CAGTGTGGTATGATGTACAGAAAGCCCTGAAGTCAACAGAGAGCAATTCACACA 773

QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThrThr 233
Db 774 AAAGGCTGTGCTCCACAGTTTGAGAAATGGCTGCAGGATAACC 816

RESULT 9
AF121344
LOCUS Mus musculus 3175 bp mRNA linear ROD 01-FEB-2000
DEFINITION Mus musculus tetraspanin Tspan-5 (Tspan5) mRNA, complete cds.
ACCESSION AF121344
VERSION AF121344.1 GI:6841032
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Garcia-Frigola, C., de Lecea, L. and Soriano, E.
JOURNAL Mouse Tspan-5 cDNA cloning
REFERENCE Unpublished
AUTHORS Garcia-Frigola, C., de Lecea, L. and Soriano, E.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1999) Dept. of Animal and Plant Cell Biology,
University of Barcelona, Av. Diagonal 645, Barcelona 08028, Spain
FEATURES
source
1..3175
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Swiss Webster/NIH"
/db_xref="taxon:10090"
gene 1..3175
/locus="Tspan5"
CDS 388..1194
/locus="Tspan5"
/feature="member of transmembrane 4 superfamily"
/codon_start=1
/product="tetraspanin Tspan-5"
/protein_id="AAF28869.1"
/db_xref="GI:6841033"
/translation="MSGKHVKGPEVSCCIKVFIFGNVIFWFLGIFLIGLWAWNEK
GVLSNISTDLGGDPVWFLVGVWFLVGGVAGGACGLAKENTFLKFFSVFGLIF
FLELTAGVLAEPVQDKWKDQLYFNNINRAIRDDIDLDQNLIDFTQEWCCGAFAD
KGNLNIYNTDSNARERCVPFSCCTQPAEDVINTQCYDARQKPEVDQIVVT
DWCVPQFKVQDLNLVAGIFIGLALQIFGLCLAQNLVSDIEAVRASW"
BASE COUNT 667 a 811 c 840 g 857 t

```

## ORIGIN

## Alignment Scores:

Pred. No.: 1..1e-71 Length: 3175  
 Score: 944.50 Matches: 178  
 Percent Similarity: 83.76% Conservative: 18  
 Best Local Similarity: 76.07% Mismatches: 34  
 Query Match: 73.62% Indels: 4  
 DB: 10 Gaps: 2

US-09-972-970-4 (1-233) x AF121344 (1-3175)

QY 1 MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysCysGlyLysTyrPhe 20  
 Db 388 ATGTCCGGGAAG-----CACTACAAAGGGTCTCTGAAGTCAGTTGTTGCAATCAATCTTC 441  
 QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40  
 Db 442 ATTTTGGCTTCAATGTCATATTTTGGTTTGGGAATAACGTTTCTTGGAAATCGGACTG 501  
 QY 41 TrpAlaTrpGlyGlyLeuSerValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60  
 Db 502 TGGGCGTGGAAATGAAAAGGTGCTCTCCAAACATCTCGTCCATCACCGACCTCGGTGGC 561  
 QY 61 LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla 80  
 Db 562 TTGTACCCAGTGTGGCTTTTCTCTGGTGGGAGGAGTGATGTTCAATCTCGGGGTTTGCA 621  
 QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100  
 Db 622 GGTGTCATCGGAGCACTTCGGGAAACACCTTTCTCTCAAGTTTTTCTGTGTCTCTG 681  
 QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTTP 120  
 Db 682 GGGATTTATTTCTTCTCGAACTACTCTCGGGGTGGCATTTGTTTCAAGACTCG 741  
 QY 121 IleArgAspGlnLeuAsnPhePheIleAsnValLeuAlaTyrArgAspAspIle 140  
 Db 742 ATCAAGACCAGCTGTATTCTTTATTAAACAACATCAGAGCCTACAGAGATGACATT 801  
 QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysCysGlyAlaArgGly 160  
 Db 802 GATCTACAGAACCTCATAGACTTCACCAGGAATATTGGCAGTCTGTGGGCTTTTGA 861  
 QY 161 ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSerArgGlu 180  
 Db 862 GCTGATGATTGGAACTTAATATTACTTCAATTCACAGATTCCTCAAGCCGAGAG 921  
 QY 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199  
 Db 922 CGATGCGGTGGCCATTTTCTGCTGCATAAAGACCCCGGAGAGATGTCTATCAACACT 981  
 QY 199 oSerValAlaMetMetSerGlySerAsnTrpSerTrpSerSerArgAlaProTyrThrPr 219  
 Db 982 CAGTGTGCTATGATGCCAGGACAGAACCAAGTTGACCAACAGATTGTAATCTACACA 1041  
 QY 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
 Db 1042 AAAGCTGTGTGCCCCAGTTTGGAGAGTGGCTACAGGACA 1081

## RESULT 10

AF053455  
 LOCUS Homo sapiens tetraspan TM4SF (TSPAN-5) gene, complete cds.  
 DEFINITION AF053455  
 ACCESSION AF053455  
 VERSION AF053455.1 GI:2995864  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1408)  
 AUTHORS Todd,S.C., Doctor,V.S. and Levy,S.

TITLE		Sequences and expression of six new members of the tetraspanin/TM4SF family	
JOURNAL	Biochim. Biophys. Acta 1399 (1), 101-104 (1998)		
MEDLINE	98390278		
PUBMED	9714763		
REFERENCE	2 (bases 1 to 1408)		
AUTHORS	Todd,S.C., Doctor,V.S. and Levy,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAR-1998) Medicine, Stanford, 300 Pasteur Dr, Stanford, CA 94305, USA		
FEATURES		Location/Qualifiers	
source	1..1408	/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
gene	/db_xref="taxon:9606"		
	352..1146		
	/gene="TSPAN-5"		
CDS	352..1146	.	
	/gene="TSPAN-5"		
	/note="transmembrane 4 superfamily"		
	/codon_start=1		
	/product="tetraspan TM4SF"		
	/protein_id="AAC69712.1"		
	/db_xref="GI:2995865"		
	/translation="MSGRHYKGPVSCIKYFPGFNVIWFVLGITFLGLGWANNEK		
	TVLNSISITDLGGFDVWLFLVVGVMFILFGAGCIGALRENTFSVFLGIFFLLEL		
	GAGLAVAFVKDITKDQYLFITNNIRAYRDIQLNLDFTQYMWCCGAFGADDNNL		
	NIYFNCITDSNARSECVPSCCTCKDPAEDVINTQCGYDARQKPEVDQQIVITKGCV		
	POPEKIQDNLNITVAGIFIGIALIIFGICLAQNLSDIEAVASW"		
BASE COUNT	309 a 354 c 384 g 345 t	16 others	
ORIGIN			
Alignment Scores:			
Pred. No.:	1.81e-70	Length:	1408
Score:	925.50	Matches:	176
Percent Similarity:	82.48%	Conservative:	17
Best Local Similarity:	75.21%	Mismatches:	33
Query Match:	72.14%	Indels:	8
DB:	9	Gaps:	3
US-09-972-970-4 (1-233) x AF053455 (1-1408)			
QY	1	MetProGlyIysHisGlnHisPheGlnGluProGluValGlyCysGlyIysTyrPhe	20
Db	352	ATGTCCGGGAAG-----CACTACAGGGTCTCGAAGTCAGTTGTTGCATCAAACTACTTC	405
QY	21	LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlalleGlyLeu	40
Db	406	ATATTGGCTTCAATGTCATATTTTGGTATTTTGGGAATAACATTTCTTGGATTGGAGCTG	465
QY	41	TrpAlaTrpGlyGluIysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly	60
Db	466	TGGCATGGAATGAAAAAGGAGTCTGTCCAAACATCTTCCATCACCGCATCTCGGGGCG	525
QY	61	LeuAspProValTrpLeuPheValValGlyGlyValMetSerValLeuGlyPheAla	80
Db	526	TTTGACCCATTTGGCTCTTCTGTGTGTGGAGGAGTGATGTTCATTTTGGATTTCGA	585
QY	81	GlyCysIleGlyAlaLeuArgGluAenThrPheLeuLeuLysPhePheSerValPheLeu	100
Db	586	GGGTGATTCGACGGCTACGGGAACACATTTTCCTT-----TCTGTGTTCTCTG	633
QY	101	GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleAlaPheValPheLysAspTrp	120
Db	634	GGAATTATTTCCTCTGGAGCTCACTGCCGGAGTCTAGCATTTGTTTCAAGACTGG	693
QY	121	IleArgAspGlnLeuAenPhePheIleAsnAsnValLysAlaTyrArgAspAspIle	140
Db	694	ATCAAGACGACGCTGATTTCCTTTATAAACAACATCAGAGCATATCGGATGACATT	753
QY	141	AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyAlaArgGly	160
Db	754	GATTTCGAAACCTCATAGACTTCACCCAGGAATATTGGCAGTGTGTCGGGCTTTTGA	813

Qy 161 ProAsnAspTrpAsnLeuAsnIleTyPheAsnCyThrAspLeuAsnProSerArgGlu 180  
 Db 814 GCTGATGATTGGAACTAAATATTACTTCAATTGCACAGATTCGAATGCAAGTCGAGAG 873

Qy 181 ArgCysGlyValProPheSerCysCysValArgAspProAla----MetSerSerThrPr 199  
 Db 874 CGATGTGGCGTTCCATTCTCTGCTGCACTAAAGATCCCGCAGAGATGTCATCAACACT 933

Qy 199 oSerValAlaMetSerCysGlySerAsnTrpSerTrpSerArgAlaProTyThrPr 219  
 Db 934 CAGTGTGGCTTATGATGCCAGGCAAAACCAAGAGTTGCACGACAGATTGTAATCTTACACG 993

Qy 219 oLysAlaValTrpAlaSerLeuArgSerGlyCysArgThr 232  
 Db 994 AAGGCTGTGTGCCCGCAGTTTGAGAGTGGTTCGACGACA 1033

RESULT 11  
 LOCUS AK024427 4445 bp mRNA linear PRI 29-SEP-2000  
 DEFINITION Homo sapiens mRNA for FLJ00016 protein, partial cds.  
 ACCESSION AK024427  
 VERSION AK024427.1 GI:10440362  
 KEYWORDS fls (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 4445)  
 AUTHORS Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.  
 TITLE The nucleotide sequence of a long cDNA clone isolated from human spleen  
 JOURNAL Published Only in DataBase (2000)  
 REFERENCE 2 (bases 1 to 4445)  
 AUTHORS Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,  
 Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba  
 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,  
 URL:http://www.kazusa.or.jp/NEDO, Tel:81-438-52-3913,  
 Fax:81-438-52-3914)  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing: Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Kazusa DNA Research Institute.  
 FEATURES  
 source  
 1. .4445  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="as00016"  
 /tissue\_type="spleen"  
 /dev\_stage="adult"  
 /note="vector:pBluescriptII SK plus"  
 1. .4445  
 /gene="FLJ00016"  
 <2041. .3090  
 /gene="FLJ00016"  
 /note="For this clone, GeneMark analysis triggered an  
 alert for spurious CDS split. The result of GeneMark  
 analysis of the nucleotide sequence of this clone is  
 accessible through http://www.kazusa.or.jp/NEDO.  
 Start codon is not identified."  
 /codon\_start=1  
 /product="FLJ00016 protein"  
 /protein\_id="BAB15717.1"  
 /db\_xref="GI:10440363"  
 /translation="GVLGGGGVDDSDSRHGGNGVARLWGHQDPPVPALKRGGLCRT  
 SGASVGGLLQHTGWPAVSATGAGPAPQVYWRKPQSLNYPCTPSLFLASLGSRGK  
 ELVSPSLAPLCHPQGWQAQGGGVATHLAVPALSQALYSCSLGFPNVRVLLGW  
 AWTWGSPLPQWSCCGAGPNDNLNLYFNCTDLPNSRRCGVPRSCCVDPDAEDVL  
 NTCQGVNRLKLELFOCGEITWKGVCQPEWLDLNLVAGVGGTALVQIETCTA

QNLVSDIKAVKANWSKWNDDPENHMLTPTISEVLSTAGPQQNSLTGAPGPAPPSRHVF  
FGLGGLYPEPTFKW"

BASE COUNT 890 a 1278 c 1321 g 956 t  
ORIGIN

Alignment Scores: 1.13e-63 Length: 4445  
Pred. No.: 853.50 Matches: 227  
Score: 25.33% Conservative: 0  
Percent Similarity: 25.33% Mismatches: 6  
Best Local Similarity: 66.52% Indels: 663  
Query Match: 9 Gaps: 2  
DB: 2

US-09-972-970-4 (1-233) x AK024427 (1-4445)

```
QY 1 MetProGlyIysHisGlnHisPheGlnGluProGluValGlyCysGlyIysTyrPhe 20
DB 120 ATGCCCGGGAAGCAGCAGATTTCCAGGAACCTGAGGTGGCTGCTGCGGGAATACTTC 179
QY 21 LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu 40
DB 180 CTGTTTGGCTTCAACATTGCTTCTGGGTGGTGGAGCCCTGTTCTGGCTATCGGCCTC 239
QY 41 TrpAlaTrpGlyIleGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly 60
DB 240 TGGGCTGGGTGAGAAGGGCTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGC 299
QY 61 LeuAspProValTrpLeuPheValValValGlyValGlyValMetSerValLeuGlyPheAla 80
DB 300 CTTCACCCCGTGGCTGTTTGTGGTAGTGGAGGCTCATGTGGTGTGGCTTTGGCT 359
QY 81 GlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerValPheLeu 100
DB 360 GGTCTGATTGGGGCTCTCGGAGAAACACCTTCTGCTCAAGTTTTTCTCCGTGTTCCTC 419
QY 101 GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp 120
DB 420 GGTCTCATCTTCTCTGGAGCTGGCAACAGGATCTCTGGCTTTGTCTTCAAGGACTGG 479
QY 121 IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle 140
DB 480 ATTGAGACCACTCAACTTCTTCATCAACAACAACAGCTCAAGGCTTACCGGACGACATT 539
QY 141 AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyr----- 152
DB 540 GACCTCCAGAACCTCATTGACTTTGCTCAGGAATACGTGAGTCCAGTGTCCAGCCTGGGA 599
QY 152 ----- 152
DB 600 CACCTGTAGGAGGCGCTTCTTGGCAGACGAATGAGGAGCAAGTTCCCTGTGTGGGAC 659
QY 152 ----- 152
DB 660 CATCTCTTACCACCTGGGCTAGCGGGCCCCCAGGAGAGAGACCAGGAGTGTCTCTGTC 719
QY 152 ----- 152
DB 720 CGAGGAGAGCAGCCCTGCTGATCCTTATGGGTCCCCATTTGCTAGTGCATAGGGCC 779
QY 152 ----- 152
DB 780 ACGGCTTTGGAAACAGAGGCAACGGGCTCCACACTAGCAGCTCCAGGCGCAGGCGCTTAC 839
QY 152 ----- 152
DB 840 GACACAGGGAGATGGAGGAATGTGACATAAAATATAGATTCTACTTCTGCTAAACC 899
QY 152 ----- 152
DB 900 TTGGSCCCACGGTATTGGCTGCAGCTATGTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 959
QY 152 ----- 152
```

```
DB 960 CTCGTGTTGCCATTGTTCCCAACACTCTGTATTATTTATATATATGTCAGCCCTTATCTCT 1019
QY 152 ----- 152
DB 1020 CATTTATATATGTTTGGCTCTGTGTATATTAGAGTTTAGACCCCTGTGGAAGGTCTG 1079
QY 152 ----- 152
DB 1080 CCTCAAACTTGCAAACTTAAACCCAGTAGAAGAGCAGCAGAGAGTTCCTGAAACCC 1139
QY 152 ----- 152
DB 1140 TATGCTCTCCACCTGCGACCATCTGCGCCACTCAGCCCTTAACACAAACACACAGTCA 1199
QY 152 ----- 152
DB 1200 ACACTGTCAACCCACAGTCACACCCACACCTAGTCACACACTCGCAGGCACACTTGATCA 1259
QY 152 ----- 152
DB 1260 ACCTACACTACACATGGTCCACAGTCACATGAGTCACACCCGATCACACCCACACA 1319
QY 152 ----- 152
DB 1320 CGGCGCAGGCGCTTGGCACCATTCACCAACAAGAGTCAGCTCTGGATCCAGAGTGAAG 1379
QY 152 ----- 152
DB 1380 CAGAGCTGTGCTGCTTCAAGCAGCTCTCAACCCAGGCGAGGAGACATAGCCAGGC 1439
QY 152 ----- 152
DB 1440 AGTTATGATGCAAGTTTCAATTCATTTCAGCACAGGATTTGTTAAGGGCTTACATATGCC 1499
QY 152 ----- 152
DB 1500 GGAACCCAGAGCCAGGCCCTGAGGATTCGGCGGGAAGAAAGCTGGGACCTGTCTCTCAG 1559
QY 152 ----- 152
DB 1560 GGCTGGCAGTCGGGAAGAGACAGATGGTGAACAAGCAGACAACAACAAATGTGTGATGTA 1619
QY 152 ----- 152
DB 1620 CATAGCTGTGTATGCAAAATTTGTACATTTGGTAAAGATCAAAATGAAGGGTGGAGAATG 1679
QY 152 ----- 152
DB 1680 GGAAGCCCTCTCGTGGAGAGCTTGAAGCCTCCCAAGGAGGAAGGAAGAGAGGGAG 1739
QY 152 ----- 152
DB 1740 AGTGTCTTAGCAGAGCAGCTGGCGCTGAGCAGGTGCCAGGAGGAGGAGGAGGCTGA 1799
QY 152 ----- 152
DB 1800 GCCCTGCCCTGGACCCAGCACATGAGGGGCATGTGGCAGGCTGACTCTCTCCAGCTCTCT 1859
QY 152 ----- 152
DB 1860 GGCTCGGCCACGGGGCTGCTGTGTCTGTGCTGCCGAGCTTGAGCTCCCGGAGGCTGTGG 1919
QY 152 ----- 152
DB 1920 GCTGCTCTCCAGCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1979
QY 152 ----- 152
DB 1980 AAGACCATGTGTTCAATGGCAGATCCAAACTCGCAATTGCTCTCTGCTTAGGGGACAGCTTA 2039
QY 152 ----- 152
DB 2040 GGGTGTCTGGTGGGGGTGATGGGGTGTGGGATGATGCCAGCCGGCATGGAGGAAATGG 2099
```



Db	766	AAAGGCTGTGTGCCCCAGCTTTGAGAAAGTGGTTCAGGACA	805
RESULT 13	AX343015		
LOCUS	Sequence 1 from Patent WO0198350.	864 bp	DNA linear PAT 12-JAN-2002
DEFINITION	AX343015		
ACCESSION	AX343015		
VERSION	AX343015.1	GI:18152215	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Reinhard, C., Jefferson, A.B., Winter, J.A. and Randazzo, F.		
TITLE	Compositions and methods for treating neoplastic disease using net		
JOURNAL	-4 modulators		
FEATURES	Patent: WO 0198350-A 1 27-DEC-2001; CHIRON CORPORATION (US)		
source	Location/Qualifiers		
	1..864		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/note="NET-4 oligonucleotide used in cell proliferation		
	assay on SW620 cells"	224 t	
BASE COUNT	162 a	229 c	249 g
ORIGIN			
Alignment Scores:			
Pred. No.:	4,71e-59	Length:	864
Score:	790.00	Matches:	139
Percent Similarity:	88.51%	Conservative:	15
Best Local Similarity:	79.89%	Mismatches:	18
Query Match:	61.57%	Indels:	2
DB:	6	Gaps:	1
US-09-972-970-4 (1-233) x AX343015 (1-864)			
QY	1	MetProGlyLysHisGlnHisPheGlnGluProGluValGlyCysGlyLysTyrPhe	20
Db	348	ATGTCCGGGAAG-----CACTACAAGGGTCTCTGTCACATCTCTCCATCACCAGATCTCGGGCGC	401
QY	21	LeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIleGlyLeu	40
Db	402	ATATTGGCTTCAAGTGCATATTTGGTTTGGGAATACATTTCTTGGAAATGGACTG	461
QY	41	TtpAlaTtpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeuGlyGly	60
Db	462	TGGGCATGGAATGAAAGAGGTTCTGTCCACATCTCTCCATCACCAGATCTCGGGCGC	521
QY	61	LeuAspProValTrpLeuPheValValGlyValMetSerValLeuGlyPheAla	80
Db	522	TTTGACCCAGTTGGCTTCTTCTGTGGTGGAGGAGTGATGTTTCATTTGGGATTGCA	581
QY	81	GlyCysIleGlyAlaLeuArgGluLeuThrPheLeuLysPhePheSerValPheLeu	100
Db	582	GGTGCATTTGGAGCCCTACGGAAACACATTTCTTCAAGTTTTTCTGTGTCTCTG	641
QY	101	GlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLysAspTrp	120
Db	642	GGAATTATTTCTTCTGAGCTCACTGCGGAGTTCTAGCATTTGTTTTCAAAGACTGG	701
QY	121	IleArgAspGlnLeuAsnPhePheIleAsnAsnValLysAlaTyrArgAspAspIle	140
Db	702	ATCAAGACCAGCTGATTTCTTTTATAACACACATCAGCAGCATATCGGGATGACATT	761
QY	141	AspLeuGlnAsnLeuIleAspPheAlaGlnGluTyrTrpSerCysGlyValaArgGly	160
Db	762	GATTTCGAAACCTCATAGACTTCACCCAGGAAATATGTCAGTCTGTGGGCTTTTGG	821
QY	161	ProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAsp	174
FEATURES			
Db	822	GCTGATGATTGGAACTAAATATTACTTTCAATTGCACAGAT	863
RESULT 14	BC025568		
LOCUS	Mus musculus DNA segment, Chr 14, ERATO Doi 226, expressed, mRNA	2428 bp	mRNA linear ROD 16-APR-2003
DEFINITION	BC025568		
ACCESSION	BC025568		
VERSION	BC025568.1	GI:19343883	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2428)		
JOURNAL	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Maruina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length		
JOURNAL	human and mouse cDNA sequences		
MEDLINE	Proc Natl Acad Sci U S A. 99 (26), 16899-16903 (2002)		
PUBMED	22388257		
REFERENCE	12477932		
AUTHORS	2 (bases 1 to 2428)		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAR-2002) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk		
	Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a>		
	Tissue Procurement: Jeffrey Green M.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: National Institutes of Health Intramural		
	Sequencing Center (NISC),		
	Gaithersburg, Maryland;		
	Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a>		
	Contact: <a href="mailto:nisc_mgc@hgrl.nih.gov">nisc_mgc@hgrl.nih.gov</a>		
	Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.B., Masiello, C., Maskeri, B., Mestrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripp, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.		
FEATURES			
	Clone distribution: MGC clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Series: IRAP Plate: 56 Row: j Column: 21		
	This clone was selected for full length sequencing because it		
	passed the following selection criteria: GenomeScan gene		
	prediction.		
	Location/Qualifiers		

Source 1. .2428 /organism="Mus musculus" /mol\_type="mRNA" /strain="FVB/N" /db\_xref="taxon:10090" /clone="MGC:36595 IMAGE:5322531" /tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal Carcinoma. 5 month old virgin mouse." /clone\_lib="NCI CGAP\_Mam6" /lab\_host="DH10B" /note="Vector: pCMV-SPORT6" 1. .2428 /gene="D14Etd226e" /note="synonym: MGC36554" /db\_xref="LocusID:52588" /db\_xref="MGI:1196325" 64. .876 /codon\_start=1 /product="tetraspanin similar to TM4SF9" /protein\_id="AAH25568.1" /db\_xref="GI:19343884" /db\_xref="LocusID:52588" /translation="MHYRYSNAEVSCHWYKLLFSNVVFLVGVVFLGVGLWVSEK GVLSDTKVRLHGDIPVVLVLMGVMTFLGFCAGCVGALRENI CLIKFPCGAI VLI FLELAVAFLEFQWVRDRFPFESNIKSYRDDIDLQNLDSLKAKNOCCGAYGPE DMDLVYFNCGASYSREKCGVPFCVCPDPAQKVNTGCGYDVRILQLSKWFDFIT KCICQALEGLPRNIYIVAGVFIALSLQLFGIFLARTLISDIEAVAGHHF" BASE COUNT 517 a 640 c 624 g 647 t ORIGIN Alignment Scores: Pred. No.: 2,86e-48 Length: 2428 Score: 671.00 Matches: 134 Percent Similarity: 66.93% Conservative: 34 Best Local Similarity: 53.39% Mismatches: 60 Query Match: 52.30% Indels: 23 DB: 10 Gaps: 5 US-09-972-970-4 (1-233) x BC025568 (1-2428) QY 2 ProGlyLysHisGlnHis-----PheGlnGluProGluValGlyCysCysGlyLys 18 Db 52 CTTCTTCTGAGATGCGATGATATATAGATACGAGCGCGAGGTGAGTCTGCTGTTACAAG 111 QY 19 TyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIle 38 Db 112 TACCTGCTCTTACGCTACAAATATGCTCTTTGGCTGGTGGAGTGTCTCTCTGAGTGC 171 QY 39 GlyLeuTrpAlaTrpGlyGlyLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeu 58 Db 172 GGGCTGTGGCATGAGCGAAAGGTGTGCTGTCGACCTCACCAGGTGACCGGTG 231 QY 59 GlyGlyLeuAspProValTrpLeuPheValValValGlyGlyValMetSerValLeuGly 78 Db 232 CATGAATCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291 QY 79 PheAlaGlyCysIleGlyAlaLeuArgGluAenThrPheLeuLysPhePheSerVal 98 Db 292 TTCGAGAGCTGTGTCGGGGCCCTCGAGAGACATCTGCTGCTCAAGTTTCTGTGGG 351 QY 99 PheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLys 118 Db 352 GCCATTGTGCTCATCTCTCTGGAACCTGCGCGTGGCGTGGCTGCTTTATTCAC 411 QY 119 AspTrpIleArgAspGlnLeuAsnPhePheIleAsnAsnValIleAlaIleArgAsp 138 Db 412 GACTGGGTGAGAGACCGGTTCGGGAATTTCTTCGAGAGACACATCAAGTCTATCGGAT 471 QY 139 AspIleAspLeuGlnAsnLeuIleAspPheAlaGlnGluTrpTrpSerCysCysGlyAla 158 Db 472 GACATCGACCTGCGAGACCTCATTCAGTCCCTTCAGAAAAGGTAAATCATGCTCGGGCT 531 QY 159 ArgGlyProAsnAspTrpAsnLeuAsnIleTyrPheAsnCysThrAspLeuAsnProSer 178

Db 532 TAGGGCCCTGAAGACTGGGACCTCAATGTCTACTTCAACTGCGAGTGGTCCACTACAGC 591 QY 179 ArgGluArgCysGlyValProPheSerCysCysValArgAspProAlaMet-SerSer-- 197 Db 592 CGAGAGAAATGTGGGTGATCCCTTCTCTGCTGTGTGCCAGATCTCTGCACAAAAGTGTG 651 QY 198 -ThrProSerValAlaMetMetSerGly----- 206 Db 652 AACACACAGTGTGGCTATGATGTCCGGATTTCAGCTGAAGAGCAAGTGGGATGATTCATC 711 QY 207 -----SerAsnTrpSerTrpSerArgAlaProTyrThrProLysAlaVal 222 Db 712 TTTACAAAAGGATGATCCAGGCTCTTGAAGGCTGGCTGCCAGGAACA---TCTACATT 768 QY 222 lTrp-----AlaSerLeuArgSerGlyCys 230 Db 769 GTGGCTGGTGTCTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 799 RESULT 15 LOCUS BC024611 DEFINITION Mus musculus DNA segment, Chr 14, ERA10 Doi 226, expressed, mRNA ACCESSION BC024611 VERSION BC024611.1 GI:19354053 KEYWORDS MGC. SOURCE Mus musculus ORGANISM Mus musculus (house mouse) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2498) Strausberg,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,P., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fanej,J., Helton,E., Kettner,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalieu,D.E., Schermer,A., Schein,J.E., Jones,S.J. and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 23388257 12477932 2 (bases 1 to 2498) Strausberg,R. Direct Submission Submitted (01-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA NIH-MGC Project URL: <http://mgc.nci.nih.gov> Contact: MGC help desk Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov) Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: <http://www.hgsc.bcm.tmc.edu/cdna/> Contact: [amgobcm.tmc.edu](mailto:amgobcm.tmc.edu)



Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK plate: 37 Row: e Column: 12  
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.

## FEATURES

```

production, similarity, but not identity
location/qualifiers
1. .2498
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db xref="taxon:10090"
/clone="MGC:28503 IMAGE:418"
/tissue_type="Salivary gland"
/clone_1fb="NCI CGAP_SG2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .2498
/gene="D14Erdc26e"
/note="synonym: MGC36554"
/db xref="LocusID:52588"
/AB_xref="MGI:1106325"

```

CDS 55.867

```

/codon_start=1
/product="tetraepanin similar to TM4SF9"
/protein_id="AAH24611.1"
/db_xref="GI:19354054"
/db_xref="LocusID:52588"
/translation="MHYRYRYSAEYSCWKYLLFSVNIYVFLAGVVFVLCGLWAWSEK
GVLSLDKTYRHLGHDPVVLVLVLMVGVMVMTLGFAGCVGALRENICLLKFGCAIVLIF
FLELAVLAFVLDQWRREFRFFESNKIYRDDIDLQNLDSLQKNCQCCGAYGPE
DRLNLYFCGSLGYSREKGVPPFSCVCPDPKQGVNTQGYDVRITQLKSKWDEFTT
KGCIOALNGLPRLNIYIVAGVFIATSLLOIFGLFIARTLSDIEAVKACGHS"

```

BASE COUNT	542 a	644 c	631 g	681 t
ORIGIN				

Alignment Scores:

Assignment Scores:			
Pred. No.:	2,966-48	Length:	2498
Score:	671.00	Matches:	134
Percent Similarity:	66.93%	Conservative:	34
Best Local Similarity:	53.39%	Mismatches:	60
Query Match:	52.30%	Indels:	23
DB:	10	Gaps:	5

US-09-972-970-4 (1-233) x BC024611 (1-2498)

QY	2	ProGlyLysHisGlnHis-----PheGlnGluProGluValGlyCysCysGlyLys	18
Db	43	CCTTCTCAGAAGATGCATATTATAGATACTCGAAAGCCGAGGTGAGTCTGGTACAAG	102
QY	19	TyrPheLeuPheGlyPheAsnIleValPheTrpValLeuGlyAlaLeuPheLeuAlaIle	38
Db	103	TACCTGCTCTTCAGCTACAATATCGTCTTTTGGCTGGAGTTGTCTTCCTTGGAGTC	162
QY	39	GlyLeuTrpAlaTrpGlyGluLysGlyValLeuSerAsnIleSerAlaLeuThrAspLeu	58
Db	163	GGCTGTGGGCATGAGCGAAAGGGTGTGTGTCCACCTCACCAAGGTGACCCGGTTG	222
QY	59	GlyGlyLeuAspProValTrpLeuPheValValGlyValMetSerValLeuGly	78
Db	223	CATGGAATCGACCGCTGCTGCTGATGGTGGCGTGGTGATGTCACACTGGGA	282
QY	79	PheAlaGlyCysIleGlyAlaLeuArgGluAsnThrPheLeuLeuLysPhePheSerVal	98
Db	283	ITTCGAGGCTGTGTGGGGGGCCCTCCGAGAGAACATCTGCCTGCTCAAGTTTTTCTGTGGG	342
QY	99	PheLeuGlyLeuIlePhePheLeuGluLeuAlaThrGlyIleLeuAlaPheValPheLys	118
Db	343	GCCAAATGTGCTCATCTCTCTCTGGAATCGGGCGTGGCGCTGTGGCTTTTATTCCAA	402

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 16:37:04 ; Search time 5185 Seconds  
(without alignments)  
11896.772 Million cell updates/sec

Title: US-09-972-970-2

Perfect score: 2538

Sequence: 1 ccacgctcgccgcagcc.....aaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	796	31.4	1201	9	AL528502
2	771.8	30.4	1194	9	AL529630
3	743.8	29.3	1494	11	AK008761
4	692.2	27.3	957	9	AL543914

5	677.8	26.7	1005	14	BY708665
6	663.8	26.2	938	13	BU856801
7	655.2	25.8	849	10	BE615772
8	632.2	24.9	1032	12	BM563474
9	622	24.5	1201	9	AL556793
10	621	24.5	1028	10	BG477727
11	619.2	24.4	901	14	CA454987
12	611.6	24.1	888	13	BX447619
13	606.4	23.9	781	12	BG770931
14	602.4	23.7	718	12	B1914325
15	568.4	22.4	902	10	BG478644
16	559.2	22.0	970	12	B1909709
17	552.2	21.8	782	12	B1158921
18	543.4	21.4	1020	12	B1557863
19	541.2	21.3	654	12	B1829529
20	533.2	21.0	912	12	B1915506
21	528.2	20.8	832	10	BF981395
22	521	20.5	594	10	BF043938
23	519.8	20.5	657	12	B1850548
24	515.8	20.3	916	13	BQ24634
25	503.6	19.8	588	10	BE615323
26	489.6	19.3	538	10	BE683087
27	489.6	19.3	960	10	BF119347
28	480.6	18.9	1108	13	BQ072716
29	478.6	18.9	685	14	BY734818
30	472.6	18.6	769	10	BF122628
31	468.2	18.4	600	12	B1987393
32	463	18.2	634	14	CB215495
33	451.8	17.8	704	10	BG422301
34	449.4	17.7	3141	11	AK013350
35	443.4	17.5	714	10	BG017161
36	443.4	17.5	883	14	CA986012
37	442	17.4	576	14	CB607991
38	440	17.3	932	14	CB321353
39	435.4	17.2	920	13	BQ941931
40	431.6	17.0	553	14	CB613921
41	430.2	17.0	933	12	BM475708
42	430	16.9	721	10	BF983242
43	425.6	16.8	753	14	CA315923
44	418.4	16.5	2108	11	AK033543
45	416.8	16.4	1116	11	AK015705

#### ALIGNMENTS

```

RESULT 1
AL528502
LOCUS      AL528502      1201 bp      mRNA      linear      EST 23-MAY-2003
DEFINITION AL528502 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            CDNA clone CSODC029YJ10 5-PRIME, mRNA sequence.
ACCESSION  AL528502
VERSION     AL528502.2  GI:31066352
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1201)
AUTHORS   Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
COMMENT    On Feb 13, 2001 this sequence version replaced gi:12791995.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3528.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSODC029DB05Qp16cluster=3528.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :

```

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DC029D05QPI.

## FEATURES

source

1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC029YJ10"  
/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT

ORIGIN

236 a 331 c 340 g 262 t 32 others

Query Match 31.4%; Score 796; DB 9; Length 1201;  
Best Local Similarity 91.3%; Pred. No. 9e-97;  
Matches 860; Conservative 6; Mismatches 71; Indels 5; Gaps 2;

QY	12	GCCGAGCCGCGGGCTAGCCCGGGCGGCTCTAGCCAGGGCGGCGCGGTGGAGGCGG	71
DB	64	GCGGAGCCGCGGGCTAGCCCGGGCGGCTCTAGCCAGGGCGGCGGCGGGCGGTG	123
QY	72	ATCCGGCC-CGCGCTCCGGTTCGCGGGCGGGCGGCTGCTACCATGCCGGGCAAG	130
DB	124	GCGCTGGCTCCCGGCTCCGGTTCGCGGGCGGGTGGCGCTCACCATGCCGGCAAG	183
QY	131	CACGAGCACTTCAGGAACCCGAGGTCGGCTGCTCGGGGAATACTTCTGTTGGCTTC	190
DB	184	CACGAGCACTTCAGGAACCTAGGTCGGCTGCTCGGGGAATACTTCTGTTGGCTTC	243
QY	191	AACATGTTTTGCGGTCTGGAGCCCTGTTCTGGCCATCGGCTCTGGCCCTGGGGT	250
DB	244	AACATGTTCTTGGGTCTGGAGCCCTGTTCTGGCTATCGGCTCTGGCCCTGGGGT	303
QY	251	GAGAGGGTGTCTCCCAACATCTGCGCTGACCGATCTGGAGGCTCGACCTGTG	310
DB	304	GAGAGGGGTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCTTGAACCCG	363
QY	311	TGGCTGTTGTAGTGTGGAGCGTCATGTCGGTCTTGGCTTGGCGGCTGCATCGGG	370
DB	364	TGGCTGTTGTAGTGTGGAGCGTCATGTCGGTCTTGGCTTGGCGGCTGCATCGGG	423
QY	371	GCTCTCCGGAGAACACTTCTGCTCAAGTTTCTCAGTGTCTTGGCCCTCATCTTC	430
DB	424	GCCCTCCGGAGAACACCTTCTGCTCAAGTTTCTCAGTGTCTTGGCCCTCATCTTC	483
QY	431	TTCTCGGAGCTGCACAGGATCTTGGCTTCTGATTTCAAGGACTGGATTGAGACCG	490
DB	484	TTCTCGGAGCTGCACAGGATCTTGGCTTCTGATTTCAAGGACTGGATTGAGACCG	543
QY	491	CTCAATTTCTTCAATTAACAAACGTCAGGCGCTATCGGGATGACATTCACCTCCAG	550
DB	544	CTCAATTTCTTCAATTAACAAACGTCAGGCGCTATCGGGATGACATTCACCTCCAG	603
QY	551	CTCATGACTTGTCTCAGGAATATGTTCTTGGTGGAGCCGAGGCGCTTAATGACTGG	610
DB	604	CTCATGACTTGTCTCAGGAATATGTTCTTGGTGGAGCCGAGGCGCTTAATGACTGG	663
QY	611	AACTTCAATATCTATTCAACTGCACTGACTTTGAACCGGAGCGGAGGCGCTCGGG	670
DB	664	AACTTCAATATCTATTCAACTGCACTGACTTTGAACCGGAGCGGAGGCGCTCGGG	723
QY	671	CCCTTCTCTGCTGTGTCAAGGACCCCTTGC- ---GATGTCTTCAACACCCAGTGTGG	726
DB	724	CCCTTCTCTGCTGTGTCAAGGACCCCTTGCAGGAGATGCTCTCAACACCCAGTGTGG	783
QY	727	GATGTCCGGCTCAAACTGAGCTGAGCAGAGGGGCTCATACACCAAAAGGCTGTGTG	786
DB	784	GACGTCCGGCTCAAACTGAGCTGAGCAGAGGGGCTCATACCAACCAAAAGGCTGTGTG	843

QY	787	GCCAGTTTGAGAACTGGCTGCAGACAACCTGATCGTGGTGGTGGGTCTTTGTGGGC	846
DB	844	GCCAGTTTGAGAACTGGCTGCAGACAACCTGATTTGGTGGCGGAGTCTTCATGGGC	903
QY	847	ATCGCTCTCCCTCCAGATCTTTGGTATCTGCTGGCCGAGAACCTTCTGATGATCAAG	906
DB	904	ATCGCTCTCCCTCCAGATCTTTGGCATCTGCTGGCCGAGAACCTTCTGATGATCAAG	963
QY	907	CGAGTGAAGGCCAACTGGATCAAAACATGATGATGGCTACAA 948	
DB	964	GSAGTGAAGGCCAACTGGAGSWAATGGATGATGATTTGAA 1005	

## RESULT 2

AL529630

LOCUS

DEFINITION

AL529630 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens

CDNA clone CS0DD005YB20 5-PRIME, mRNA sequence.

ACCESSION

AL529630

VERSION

AL529630.2

GI:31067473

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1194)

Li.W.B., Gruber,C., Jessee,J. and Polayes,D.

Full-length cDNA libraries and normalization

Unpublished

On Feb 13, 2001 this sequence version replaced gi:12793123.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

3528.f For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DD005DA100P1&amp;cluster=3528.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DD005DA100P1.

Location/Qualifiers

1. .1194

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DD005YB20"

/tissue\_type="NEUROBLASTOMA COT 50-NORMALIZED"

/clone\_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT

ORIGIN

237 a 331 c 336 g 253 t 37 others

Query Match 30.4%; Score 771.8; DB 9; Length 1194;

Best Local Similarity 86.0%; Pred. No. 1.5e-93;

Matches 923; Conservative 11; Mismatches 104; Indels 35; Gaps 6;

QY

12

GCCGAGCCGCGGGCTAGGCCCCGGCGGCTCTAGCCAGGGCGGCGCGGTGGAGGCGG

71

DB

54

GCCGAGCCGCGGGCTAGGCCCCGGCGGCTCTAGCCAGGGCGGCGGCGCGGTG

113

QY

72

ATCCGGCC-CGCGCTCCGGTTCGCGGGCGGCTCCCGGAGGCTGCTCACCATGCCGGCAAG

130

DB

114

GCGCTGGCTCCCGGCTCCGGTTCGCGGCCGCGGCTGCGGCTCACCATGCCGGCAAG

173

QY

131

CACCAGCACTTCCAGGAACCCGAGTCCGCTGCTCGCGGAAATACCTCTCTGTTGGCTTC

190

DB

174

CACCAGCACTTCCAGGAACCTGAGTCCGCTGCTCGCGGAAATACCTCTCTGTTGGCTTC

233





more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1005AG04QPI&cluster=3528.f>. Contact :  
 Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0D1005AG04QPI.  
 Location/Qualifiers  
 1. .957  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1005Y07"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /note="First strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with NotI and cloned into the NotI and EcoRV  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## FEATURES

source

BASE COUNT	176 a	279 c	288 g	205 t	9 others
ORIGIN					
Query Match	27.3% Score 692.2; DB 9; Length 957;				
Best Local Similarity	87.9%; Pred. No. 6.3e-83;				
Matches	813; Conservative	3; Mismatches	66; Indels	43; Gaps	4;
Qy	12	GGCCAGCCCGCGGCTAGGCCCCCGCGCGGCTTAGCCAGGCGCGCGCGCGCGCGGCGG	71		
Db	64	GGCCAGCCCGCGGCTAGGCCCCCGCGCGGCTTAGCCAGGCGCGCGCGCGCGGCGG	123		
Qy	72	ATCCGCGC-CCGCGCTCCGCTCCGCGCGCGCGCGCGGCGGCTGCTCACCATGCGCGGCAAG	130		
Db	124	GGCGTGGCTCCGCGCTCCGCTCCGCGCGCGCGCGGCGGCTGCTCACCATGCGCGGCAAG	183		
Qy	131	CACAGACATTCACAGAACCCGAGGTCGGCTGCTGCGGGAATCTCTCTGTTGGCTTC	190		
Db	184	CACAGACATTCACAGAACCTGAGTGGCTGCTGCGGGAATCTCTCTGTTGGCTTC	243		
Qy	191	AACATTGTTTCTGGGTGCTGGAGCCCTGTTCTGCGCATCGCGCTCTGGGCTCGGGT	250		
Db	244	AACATTGTTTCTGGGTGCTGGAGCCCTGTTCTGCGCATCGCGCTCTGGGCTCGGGT	303		
Qy	251	GAGAAGGTTTCTTCCACATCTTCGCTGACCGATCTGGAGGCTTCGACCTGTTG	310		
Db	304	GAGAAGGCTTCTTCGACATCTCAGCGCTGACAGATCTGGAGGCTTCGACCTGTTG	363		
Qy	311	TGCGTGTTCGTAGTGTGGAGCGTCATCGCTGCTGGCTTCGCGCTCGACGCGG	370		
Db	364	TGCGTGTTCGTAGTGTGGAGCGTCATCGCTGCTGGCTTCGCGCTCGACGCGG	423		
Qy	371	GCTCTCCGGAGAACACTTTCCTGCTCAAGTTTTCCTCAGTGTTCCTTGGCCTCATCTTC	430		
Db	424	GCCTCCGGGAGAACCTTTCCTGCTCAAGTTT-----	456		
Qy	431	TTCTGGAGTGGCAACAGGATCTTGGCTTCGTTTCAAGACTGGATTGAGACCCAG	490		
Db	457	-----CTGGCAACAGGATCTTGGCTTCGTTTCAAGACTGGATTGAGACCCAG	507		
Qy	491	CTCAATTTCTTCAATTAACACACGTCAGGCTTACGGATGACATTCACCTCCAGAAC	550		
Db	508	CTCAACTTTCTATCAACACACGTCAGGCTTACGGATGACATTCACCTCCAGAAC	567		
Qy	551	CTCATTGATTTGCTCAGGAATAT--TGCTTGTGCTGCGAGCGCCGAGGCGCTTAATGACT	608		
Db	568	CTCATTGATTTGCTCATGTGTTTCTGTTTGTGCTGCGAGSCCGAGGCGCTCATGACT	627		
Qy	609	GGAACTCAATATCTATTTCACTGACATCTTGAAACCCGAGCGGAGCGGCGG	668		
Db	628	GGAACTCAATATCTATTTCACTGACATCTTGAAACCCGAGCGGAGCGGCGG	687		
Qy	669	TGCGCTTCTCTGCTGTGTCAGGACCCCTGC---GATGCTCTCAACCCAGTGTGCT	724		
Db	688	TGCGCTTCTCTGCTGTGTCAGGACCCCTGCAGGAGATGCTCTCAACCCAGTGTGCT	747		

Qy	725	ATGATGTCCGCGCTCAAACTGGAGCTGGAGCAGGCGCTCCATACACACCAAGGCTGTG	784
Db	748	ACGAGCTCCGCGCTCAAACTGGAGCTGGAGCAGGCGCTTCATCCACCAAGGCTGTG	807
Qy	785	TGGGCCAGTTTGGAGAGTGGCTGCAGGACCAACCTGATCGTGGCTGGGCTCTTTGTGG	844
Db	808	TGGGCCAGTTTGGAGAGTGGCTGCAGGACCAACCTGATTTGGTGGCGGAGTCTTCATGG	867
Qy	845	GCATCGCTCTCTCCAGATCTTTGGTATCTGCTTGGCCAGAACCTTGTGAGTGACATCA	904
Db	868	GCATCGCTCTCTCCAGATCTTTGGCATCTGCTTGGCCAGAACCTTGTGAGTGACATCA	927
Qy	905	AGCAGTGAAGGCAACTGGATCAA	929
Db	928	AGCAGTGAAGGCAACTGGATCAA	952
RESULT 5			
BY708665			
LOCUS			
DEFINITION	BY708665 RIKEN full-length enriched, adult male stomach Mus		
ACCESSION	musculus cDNA clone 2210021G21 5', mRNA sequence.		
VERSION	BY708665		
KEYWORDS	BY708665.1 GI:27119857		
SOURCE	EST.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1005)		
	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reid, J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyszynski, B. S., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.		
TITLE	Analysis of the mouse transcriptome based on functional annotation		
JOURNAL	of 60,770 full-length cDNAs		
MEDLINE	Nature 420, 563-573 (2002)		
PUBMED	22354683		
COMMENT	12466851		
	Contact: Yoshihide Hayashizaki		
	Laboratory for Genome Exploration Research Group, RIKEN Genomic		
	Sciences Center (GSC), Yokohama Institute		
	The Institute of Physical and Chemical Research (RIKEN)		
	1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan		
	Tel: 81-45-503-9222		
	Fax: 81-45-503-9216		
	Email: genome-res@gsr.riken.go.jp,		
	URL: <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a>		
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Hirozane, T., Imotani, K., S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno		





```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6646576"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/notes="Organ: breast; Vector: pOTB7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
```

```
BASE COUNT      162 a   290 c   279 g   207 t
ORIGIN
Query Match      26.2%; Score 663.8; DB 13; Length 938;
Best Local Similarity 91.4%; Pred. No. 3.8e-79;
Matches 738; Conservative 0; Mismatches 62; Indels 7; Gaps 3;
QY 12 GCGGAGCGCGGGCTAGGCCCCGGCGGCTCTAGCCAGGCGCGCGCTGGAGGGCGG 71
Db 5 GCGGAGCGCGCGGGCTAGGCCCCGGCGGCTCTAGCCAGGCGCGCGCTGGAGGGCGG 64
QY 72 ATCCCGGC-CCCGGCTCCGGTTCGCCGGCGCGGGCGGCTGTACCATGCGCGGCAAG 130
Db 65 GGCCTGGCTCCCGGCTCCGGTTCGCCGGCGCGGGTGGCGGTCAACATGCCCGGCAAG 124
QY 131 CACCAGCACTTCAGGAACCCGAGGTCCGGTCTGCGGGAAATPACTCTCTGTTGGCTTC 190
Db 125 CACCAGCACTTCAGGAACCTGAGTCCGGTCTGCGGGAAATPACTCTCTGTTGGCTTC 184
QY 191 AACATTGTTTCTGGGTGTGGAGCCCTGTTCTGGCCATCGCCCTCTGGGCTGGGGT 250
Db 185 AACATTGTTTCTGGGTGTGGAGCCCTGTTCTGGCTATCGGCCCTCTGGGCTGGGGT 244
QY 251 GAGAGGGTGTCTCTCAACATCTCTGCGCTGACCGATCTGGAGGCGCTCGACCTGTG 310
Db 245 GAGAGGGGTCTCTCTCAACATCTCAGCGCTGACAGATCTGGAGGCGCTTGACCCCGTG 304
QY 311 TGGCTGTTTGTAGTGTGGAGCGTCAATGTCGTGCTGGGCTTTCGGGCTGCATCGGG 370
Db 305 TGGCTGTTTGTAGTGTGGAGCGTCAATGTCGTGCTGGGCTTTCGGCTGCATCGGG 364
QY 371 GCTCTCCGGAGAACCTTCTGCTCAAGTCTTCTAGTGTCTCTGGGCTCTGAGCTTC 430
Db 365 GGCCTCCGGGAGAACCTTCTGCTCAAGTCTTCTGCTGTTCTCGGCTCTCATCTTC 424
QY 431 TTCTGGAGCTGGCAACAGGATCTTGGGCTTCTGATTCAGGACTGGATTTCGAGACCA 490
Db 425 TTCTGGAGCTGGCAACAGGATCTTGGGCTTCTGATTCAGGACTGGATTTCGAGACCA 484
QY 491 CTCATATTCCTATTAACAAACAGTCAAGGCTATTCGGATGACATTGACCTCCAGAAC 550
Db 485 CTCATATTCCTATTAACAAACAGTCAAGGCTATTCGGGAGCAGACATTGACCTCCAGAAC 544
QY 551 CTCATGACTTTGCTCAGGAATATTTGGTCTTCTGCGGAGCCGAGGCGCTAATGACTGG 610
Db 545 CTCATGACTTTGCTCAGGAATATTTGGTCTTCTGCGGAGCCGAGGCGCCCAATGACTGG 604
QY 611 AACCTCAATATCTATTCACTGCACTGATTTGAAACCGAGCGGAGGCGCTGGGGTG 670
Db 605 AACCTCAATATCTATTCACTGCACTGATTTGAAACCGAGCGGAGGCGCTGGGGTG 664
QY 671 CCCTTCTCTGCTGTGTAGGAACCTTC- - -GATGTCTCTCAACACCCAGTGTGGCTAT 726
Db 665 CCCTTCTCTGCTGTGTAGGAACCTTCGGGAGGATGTCTCTCAACCCAGTGTGGCTAT 724
QY 727 GATGTCCGGCTCAAACTGGAGCTGAGCAGAGGCTCATACACCAAAAGGCTGTGTG 786
Db 725 GACGTCCGGCTCAAACTGGAGCTGAGCAGCAAGGCTTCATCCACCAAAAGGCTGCGGG 784
```

```
QY 787 GGGC--AGTTTGAGAGTGGCTCCAGG 811
Db 785 GGGCCCAATTGAAAAAGTGGCTGCAAG 811

RESULT 7
BE615772
LOCUS      BE615772
DEFINITION BE615772 Homo sapiens CDNA clone IMAGE:3621871 5',
mRNA sequence.
ACCESSION BE615772
VERSION    BE615772.1 GI:9897371
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM296 row: e column: 08
            High quality sequence stop: 699.
            Location/Qualifiers
                1..849
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3621871"
                /tissue_type="adenocarcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_39"
                /note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
                Site:2: EcoRI; CDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed
                by Ling Hong in the laboratory of Gerald M. Rubin
                (University of California, Berkeley) using ZAP-CDNA
                synthesis kit (Stratagene) and Superscript II RT (Life
                Technologies)."
```

```
BASE COUNT      165 a   241 c   243 g   199 t   1 others
ORIGIN
Query Match      25.8%; Score 655.2; DB 10; Length 849;
Best Local Similarity 89.7%; Pred. No. 5.5e-78;
Matches 726; Conservative 0; Mismatches 79; Indels 4; Gaps 2;
QY 87 CCGGTTCCGGGCGCGCGGGTGTCTCACCATCGGGCAAGCACCAGCATTCACAG 146
Db 1 CCGGTTCCGGGCGCGGGTGTCTCACCATCGGGCAAGCACCAGCATTCACAG 60
QY 147 AACCCGAGGTGGCTGTCTGCGGAAATACTTCTCTTTGGCTTCAACATGTTTCTGG 206
Db 61 AACCTGAGGTGGCTGTCTGCGGAAA-NCATTCTCTTTGGCTTCAACATGTTCTTGG 119
QY 207 TGCTGGGAGCCCTGTCTGCGCATCGGCTCTGGGCTGGGGTCGAGGGGTCTCT 266
Db 120 TGCTGGGAGCCCTGTCTGCGCTATTCGGCTCTGGGCTGGGGTCGAGGGGTCTCT 179
QY 267 CCAACATCTCTGCGTGACCGCATCTGGGAGGCGCTCGACCTGTGTGGCTGTTGTAGTGG 326
Db 180 CGAACATCTCAGCGCTGACAGATCTGGGAGGCGCTTGACCCCGTGTGGCTGTTGTGGTAG 239
```

327	Qy	TTGAGGGCGTCAATGTCGTGTGGCTTTGGCGGTGCATCGGGGCTCTCCGGGAGAA	386
240	Db	TTGAGGGCGTCAATGTCGTGTGGCTTTGGCTGATTTGGGGCCCTCCGGGAGAA	299
387	Qy	CTTTCTTCGTCAAGATTTTCTCAGGTGTCCTTGGGCTCATCTTCTCCTGGAGCTGGAA	446
300	Db	CTTCTTCGTCAAGATTTTCTCAGGTGTCCTCGGTCTCATCTTCTTCTGGAGCTGGCA	359
447	Qy	CAGGGATCTTGGCCCTTCGTATTCAAGGACTGGAATTCGAGACCAGCTCAATTTCTTCAATTA	506
360	Db	CAGGGATCTTGGCCCTTTGTCTTCAAGGACTGGAATTCGAGACCAGCTCAACCTCTTTCATCA	419
507	Qy	ACAACAACGTCGAAGCCCTATCGGGATGACATTGACTCCAGAACCTCATTTGACTTTGCTC	566
420	Db	ACAACAACGTCGAAGCCCTATCCGGGACGACATTGACTCCAGAACCTCATTTGACTTTGCTC	479
567	Qy	AGGAATATTGCTCTTGTCTCGGAGCCGAGGGCCCTAATGACTTGGAACTCTCAATATCTATT	626
480	Db	AGGAATATTGCTCTTGTCTCGGAGCCGAGGGCCCAATGACTTGGAACTCTCAATATCTACT	539
627	Qy	TCAACTGCACGTGACTTGAACCCGAGCCGAGAGCGCTGGCGGGTGCCCTTCTCCTGCTGTG	686
540	Db	TCAACTGCACGTGACTTGAACCCGAGCCGAGAGCGCTGGCGGGTGCCCTTCTCCTGCTGCG	599
687	Qy	TCAGGGACCCCTGCGAGT--TCCTCAAACCCAGGTGTGGCTATGATGTCGGCTCAAACT	743
600	Db	TCAGGGACCCCTGCGAGGATGTCTCAAACCCAGGTGTGGCTACGAGCTCGGCTCAAACT	659
744	Qy	GGAGCTGGAGCAGGAGGCTCCATACACACCAAGGCTGTGTGGGCCAGTTTGAGAGTG	803
660	Db	GGAGCTGGAGCAGGAGGCTTCATCCACCAAAAGGCTGCGTGGGCCAGTCTGAGAAAGTG	719
804	Qy	GCTCGAGACAACTTGATCGTGGTGGCTCTTTTGTGGGCATCGCTCTCCTCCAGAT	863
720	Db	GCTGGCAGAAACCTGATTTGTTGGTGGCAGGAATCTTCATGGGCATGGCATCCTCCAGAT	779
864	Qy	CTTTGGTATCTGCCTGGCCCCAGAACCTTG	892
780	Db	CCTCGGATATTTCTGGCCAAAAAATCTG	808

RESULT 8				
BM563474				
LOCUS	BM563474	1032 bp	mRNA	linear
DEFINITION	AGENCOURT_6589677 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441687 5', mRNA sequence.			EST 20-FEB-2002

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1032)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 COMMENT

Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM1916 row: f column: 24  
 High quality sequence stop: 697.

**FEATURES**  
**source**

```

/clone="IMAGE:5441687"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resilient)"
/clone_lib="NIH MGC 98"
/notes="Organ: brain; Vector: pOT87; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGACAGG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: This is a NIH_MGC Library."
BASE COUNT      203 a   309 c   286 g   232 t   2 others
ORIGIN
Query Match      24.9%; Score 632.2; DB 12; Length 1032;
Best Local Similarity 82.5%; Pred. No. 5.6e-75;
Matches 804; Conservative 0; Mismatches 73; Indels 98; Gaps 3;

Qy 80 CCGCGTCCGGTTC CCGGCGCGGCGCGCGTCTCACCATTGCGGGGAGACACAGCAC 139
Db 17 CCGCGTCCGGTTC CCGGCGCGGCGGCGTCCACCATGCCGCGAAGCACACAGCAT 76
Qy 140 TTCAGGAACCCAGAGTCGGCTGCTCGGGGAAATACTTCTGTTGGCTTCAACATGTGT 199
Db 77 TTCAGGAACCCAGAGTCGGCTGCTCGGGGAAATACTTCTGTTGGCTTCAACATGTGT 136
Qy 200 TTCTGGGTGCTGGAGCGCTGTCCTGGCCATCGGCGCTCTGGGCGTGGGTGAGAAGGGT 259
Db 137 TTCTGGGTGCTGGAGCGCTGTCCTGGCTATCGGCGCTCTGGGCGTGGGTGAGAAGGGC 196
Qy 260 GTTCTCTCCAACATCTCTGCGTCAGCATCTGGGAGGCTCGAACCTGTGTGGCTGTTT 319
Db 197 GTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCTTGACCCGCTGTGGCTGTTT 256
Qy 320 GTAGTGGTTGGAGGCGTCATGTCGGTGTGGGCTTTGCGGCTGCAATCGGGGCTCTCCGG 379
Db 257 GTGGTAGTTGGAGGCGTCATGTCGGTGTGGGCTTTGCTGGCTGCAATGGGGCGCTCCGG 316
Qy 380 GAGAACACATTTCTGCTCAAGTTTTTCTCAGTGTCTTGGCGCTCATCTTCTCTCGGAG 439
Db 317 GAGAACACATTTCTGCTCAAGTTTTT
Qy 440 CTGGCAACAGGAGATCTTGGCGTTTCGATTTCAAGGACTGGATTGAGACACGAGCTCAATTTTC 499
Db 341 -----GACTGGATTGAGACACGAGCTCAACCTC 367
Qy 500 TTCATTAAACAACAGTCGAAGGCTATCGGGATGACATTTGAATCTCCAGAACCTCATGTGAC 559
Db 368 TTCATCAACAACAAGTCGAAGGCTACTACGGGACGACATTTGACCTCCAGAACCTCATGTGAC 427
Qy 560 TTGTGCTCAGGAATATTTGCTTGTCTCGGAGCGCCGAGGGCGCTTAATGACTGGAACTCTCAAT 619
Db 428 TTGTGCTCAGGAATATCTGGTCTTGTCTCGGAGCGCCGAGGCGCCCAATGACTGGAACTCTCAAT 487
Qy 620 ATCTATTTCAACTGCACTGACTTTGAACCCGAGCGGAGAGCGCTGCGGGGTGCCCTTCTCC 679
Db 488 ATCTACTTCAACTGCACTGACTTTGAACCCGAGCGGAGGCGCTGCGGGGTGCCCTTCTCC 547
Qy 680 TGTGTGTGAGGAACCTTCG ---GATGTCCTCAACACCCAGTGTGGCTATGATGCCGG 735
Db 548 TGTGTGTGAGGAACCTTCGAGGAGTGTCTCAACACCCAGTGTGGCTATGATGCCGG 607
Qy 736 CTCAAACTCGAGCTGAGCAGCAGGGCTCCATACACCAAGGCGTGTGGGCGAGTTT 795
Db 608 CTCAAACTCGAGCTGAGCAGCAGGGCTTCATCCACCAAGGCGTGTGGGCGAGTTT 667
Qy 796 GAGAAGTGGCTGCAGGACAACTGATCGTGGTGGCTGGGGTCTTTGTGGGCAATCGCTCTC 855
Db 668 GAGAAGTGGCTGCAGGACAACTGATTTGGTGGGCGGAGTCTTATGGGCAATCGGCTCTC 727
Qy 856 CTCAGATCTTTGGTATCTGCTGGCGCCAGAACTTTGTGAGTGTGACATCAAGGCGAGTGAAG 915

```

```

Db      728  CTCAGATCTTGGCATCTGCTGGCCAGAACCTGCTGAGTGACATCAAGGCGAGTGA 787
Qy      916  GCCAATCGGATCAAAATGATGATGCTGCTCAAACTACTCAAAATCAAAACCTTGAAA 975
Db      788  GCCAATCGGAGCAATGGAATGATG-----ACTTTGAAA 822
Qy      976  CCACTGGCTTACGCCACCACTCTCAGAGGTTCCATGGCGCGCAGGGGCTCAGCGGTGCC 1035
Db      823  CCACTGGCTTACGCCACCACTTTTGAAGTCTCTGTCACGCGCGGGGCTCAACAAATCT 882
Qy      1036  TCTGCTGGGGCCCC 1050
Db      883  TCTGACTGGGGCCCC 897

RESULT 9
AL556793
LOCUS   1201 bp mRNA linear EST 31-MAY-2003
DEFINITION Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
CDNA clone CS0DK005YH21 5-PRIME, mRNA sequence.
ACCESSION AL556793
VERSION   2
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1 (bases 1 to 1201)
JOURNAL   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT   Full-length cDNA libraries and normalization
          Unpublished
          On Feb 15, 2001 this sequence version replaced gi:12899797.
          Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. This sequence belongs to sequence cluster 3528.f For
          more information about this cluster, see
          http://www.genoscope.cns.fr/
          cgi-bin/cluster.cgi?seq=CS0DK005CD11Q1&cluster=3528.f. Contact :
          Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CS0DK005CD11Q1.
FEATURES
    Location/Qualifiers
        1..1201
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DK005YH21"
            /cell_type="HELA CELLS COT 25-NORMALIZED"
            /cell_line="HELA"
            /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
            /notes="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 231 a 322 c 324 g 258 t 66 others
ORIGIN
Query Match 24.5%; Score 622; DB 9; Length 1201;
Best Local Similarity 90.9%; Pred. No. 1.2e-73;
Matches 691; Conservative 4; Mismatches 59; Indels 6; Gaps 3;
Qy      12  GCGCAGCGCGGCTAGCGCCCGCGCGCTCTAGCCAGCGCGCGCGCGCGCGCGCG 71
Db      160  GCGCAGCGCGCGGCTAGCGCCCGCGCGCTCTAGCCAGCGCGCGCGCGCGCGCG 219
Qy      72  ATCCCGCGC-CCCGGCTCCCGTTCCCGCGCGCGCGCGCGCTGTCCACATGCGGCG 130
Db      220  GCGCTGGCTCCCGGCTCCCGTTCCCGCGCGCGCGCGCGCTGTCCACATGCGGCG 279
Qy      131  CACCAGCACTTCCAGGAACCCGAGGTGCGCTGCTGCGGAAATACTTCTCTTTGGCTTC 190

```

```

Db      280  CACAGCATTTCCAGGAACCTGAGTGGCTGCTGCGGAAATACTTCTGTTGGCTTC 339
Qy      191  AACATTGTTTCTGGGTGCTGGAGCCCTGTTCTTGCCCATCGGCTCTTGCGCTGGGT 250
Db      340  AACATTGCTCTTCTGGGTGCTGGAGCCCTGTTCTTGCTATCGGCTCTTGCGCTGGGT 399
Qy      251  GAGAAGGGTGTCTCTCCACATCTCTGCGCTGACCGATCTGGGAGGCTCGACCTGTG 310
Db      400  GAGAAGGGGCTTCTCTCGAACATCTCAGCGCTGACAGATCTGGAGAGCTTGACCCCG 459
Qy      311  TGGCTGTTGTAGTGTGGAGCGCTCATGTCGTGCTGGGCTTTGGCGGCTGTCATCGG 370
Db      460  TGGCTGTTGTAGTGTGGAGCGCTCATGTCGTGCTGGGCTTTGGCTGCTGATTTGG 519
Qy      371  GCTCTCGGAGAACACTTTCTGCTCAAAGTTTCTCAGTGTTCCTTGGCTCATCTTC 430
Db      520  GCCCTCGGAGAACACTTCTGCTCAAAGTTTCTCAGTGTTCCTTGGCTCATCTTC 579
Qy      431  TTCTCGAGCTGGCAACAGGATCTTGGCTTCTGATTTCAAGGACTGGATTGAGACAG 490
Db      580  TTCTCGAGCTGGCAACAGGATCTTGGCTTCTGATTTCAAGGACTGGATTGAGACAG 639
Qy      491  CTCAATTTCTTCAATTAACAACAGCTCAAGGCTTATCGGATGACATTTGACCTCCAG 550
Db      640  CTCNACTCTTCAATCAACAACAGCTCAAGGCTTATCGGATGACATTTGACCTCCAG 699
Qy      551  CTCATTGACTTTGCTCAGGAATATTGTTCTGCTGGAGCCCGAGGCTTAATGACTGG 610
Db      700  CTCATTGACTTTGCTCAGGAATATTGTTCTGCTGGAG-SCCGAGGCCCAATGACTGG 758
Qy      611  AACCTCAATATCTATTTCACACTGCTGACCTGAAACCGAGCGCGAGCGCTGCGGGT 670
Db      759  AACCTCAATATCTATTTCACACTGCTGACCTGAAACCGAGCGCGAGCGCTGCGGGT 818
Qy      671  CCCTTCTCTGCTGCTGCTGAGGACCTGCG- - -GATGCTCAACACCGAGTGGGTAT 726
Db      819  CCCTTCTCTGCTGCTGCTGAGGACCTGCGGAGGATCTCTCAACCCAGTGTGGCTAC 878
Qy      727  GATGTCGCGCTCAAACTGGAGCTGGAGCGAGCGAGCGGCTCCA 766
Db      879  GAGTCCGCGCTCAAACTGGAGCGGAGGAGCTKGASAGCA 918

RESULT 10
BG477727
LOCUS   602521422F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:463997 5',
DEFINITION mRNA sequence.
ACCESSION BG477727
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgaabs-r@mail.nih.gov
          Tissue Procurement: ATCC/DCTD/DTF
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1404 row: k column: 06
          High quality sequence stop: 734.
          Location/Qualifiers
              1..1028
                  source

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4639997"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_20"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      194 a   296 c   310 g   228 t
ORIGIN
Query Match      24.5%; Score 621; DB 10; Length 1028;
Best Local Similarity 86.2%; Pred. No. 1.7e-73;
Matches 780; Conservative 0; Mismatches 115; Indels 10; Gaps 8;

QY 12 GCCGACGCCCGGGCTAGCCCGCGGCTCTAGCCCGGCGGCGCCCTGGAGGGCCG 71
DB 28 GCCGACGCCCGGGCTAGCCCGCGGCGGCTCTAGCCCA-GCGGCGCGGGCGCTGG 86
QY 72 ATCCGCGCCCGGCTCCGGTTCGCGGCGGCGGCGGCTGTCTACCATGCCGGCAAGC 131
DB 87 GCCTGGCTCCCGGCTCCGGTTCGCGGCGGCGGCGGCTCACCATGCCGGCAAGC 146
QY 132 ACCAGCACTTCCAGGAACCCGAGTCCGCTGTCTGCGGGAATACTTCTCTTTGGCTTCA 191
DB 147 ACCAGCACTTCCAGGAACCTGAGTCCGCTGTCTGCGGGAATACTTCTCTTTGGCTTCA 206
QY 192 ACATTGTTTCTGGTCTGGAGCCCTGTCTCTGCGCATCGGCTCTGGGCTCTGGGGTG 251
DB 207 ACATTGTTTCTGGTCTGGAGCCCTGTCTCTGCGCATCGGCTCTGGGCTCTGGGGTG 266
QY 252 AGAAGGGTGTCTCTCAACATCTCTGCGGTGACCGATCTGGAGGCTCGACCTGTGT 311
DB 267 AGAAGGGTGTCTCTCGAATCTCTGCGGTGACCGATCTGGAGGCTCTGACCCGTGT 326
QY 312 GCCTGTTTGTAGTGGAGGGCTCATGTCCGCTGTGGCTTTCGCGGCTGATCGGG 371
DB 327 GCCTGTTTGTAGTGGAGGGCTCATGTCCGCTGTGGCTTTCGCGGCTGATCGGG 386
QY 372 CTCTCCGGAGAACACTTCTCTCAAGTTTCTCAGTGTTCTTGGCTCATCTTCT 431
DB 387 CCCTCCGGAGAACACCTTCTCTCAAGTTTCTCAGTGTTCTTGGCTCATCTTCT 446
QY 432 TCCTGGAGCTGGCAACAGGATCTTGGCTTCTGATTCAGGACTGGATTCGAGACCAGC 491
DB 447 TCCTGGAGCTGGCAACAGGATCTTGGCTTCTGATTCAGGACTGGATTCGAGACCAGC 506
QY 492 TCAATTTCTTCAATTAACAACAGCTCAAGGCTTATCGGATGACATTTGACCTCCAGAAC 551
DB 507 TCAACCTTTTCAATTAACAACAGCTCAAGGCTTATCGGATGACATTTGACCTCCAGAAC 566
QY 552 TCATTGACTTGTCTCAGGAATATTGTTCTGCTGCGAGCCCGAGGCGCTTAATGACTGGA 611
DB 567 TCATTGACTTGTCTCAGGAATATTGTTCTGCTGCGAGCCGA-GGCCCCAATGACTGGA 625
QY 612 ACCTCAATATCTATTTTCAAC-TGCACCTGACTTGAACCCGAGCCGAGAGCGCTCGGGGTG 670
DB 626 ACCTCAATATCTACTTCACTTGCATGACTTGAACCCGAGGAGGCGCTCGGGGTG 685
QY 671 CCTTCTCTCTGCTGTGTAGGGAC---CTGCGATGTCTCTCAACACCAAGTGT-GGCTAT 726
DB 686 GCCTTCTCTCTGCTGTGTAGGACCTGCGGAGGATGTCTCTCAAAACCCAGTGTGGCTAC 745
QY 727 GATGTCGGGCTCAAACT-GGAGTGTGAGCAGCAGGAGGCTCCATACACCAAGAGGCTGTGT 785
DB 746 AACGTCGGGCTCAAACTGGAGAGCTGGAGCAGCAGGAGGTTTATCATCACCAAAAGG-TGCGT 804

```

```

QY 786 GGGCCAGTTTGAGAGTGGCTGCAGGACACCTGATCGTGGTGGCTGGGCTTTTGTGG 845
DB 805 GGGCCAGTTTCAAAAGTGGCTTGCAGGACACCTGATTTGT-GTGGGAATCTTTTCATTTGG 863
QY 846 CATCGCTCTCTCCAGATCTTTGGTATCTGCTGGCCAGAACCTTGTGAGTGACATCAA 905
DB 864 AATCGACCTCTCAAGATTTTGGGATCTGGCAGGGGCACAACTTCGTAAGGAACACAA 923
QY 906 GGCAG 910
DB 924 GAGAG 928

CA454987      901 bp      mRNA      linear      EST 12-NOV-2002
AGENCOURT 10735735 MAPcL Homo sapiens cDNA clone IMAGE:6722499 5',
mRNA sequence.
CA454987
CA454987.1 GI:24905277
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Kristi A. Eglund, Ira Pastan
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14285 row: f column: 03
High quality sequence stop: 683.

FEATURES             Location/Qualifiers
     source            1..901
     organism="Homo sapiens"
     mol_type="mRNA"
     db_xref="taxon:9606"
     clone="IMAGE:6722499"
     cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HM1
     , LNCaP"
     lab_host="EMDH10B"
     clone_lib="MAPcL"
     notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
     Subtracted with brain, liver, lung, kidney and muscle.
     Directionally cloned. Priming method: oligo-dT. Average
     insert size: 1800 bp. Library amplification: 26,000 fold.
     Kristi A. Eglund, James J. Vincent, Robert Strausberg,
     Bungkok Lee & Ira Pastan: Discovery of new breast
     cancer genes encoding membrane and secreted proteins.
     Manuscript submitted."
BASE COUNT      157 a   275 c   269 g   200 t
ORIGIN

Query Match      24.4%; Score 619.2; DB 14; Length 901;
Best Local Similarity 92.1%; Pred. No. 3.2e-73;
Matches 687; Conservative 0; Mismatches 53; Indels 6; Gaps 3;

QY 12 GCGCAGCGCGGGCTAGGCCCCGGCGGCTCTAGCCAGCGGCGCGCGCGCGGCGCG 71
DB 25 GCGCAGCGCGCGGGCTAGGCCCCGGCGGCTCTAGCCAGCGGCGCGCGCGCGCGCTG 84
QY 72 ATCCCGGC-CCCGGCTCCGGTTCCTCCCGGCGGCGGCGGCTGTCTACCATGCCGGGCAAG 130
DB 85 GGCCTGGCTCCCGGCTCCGGTTCCTCCCGGCGGCGGCGGCTGTCTACCATGCCCGGCAAG 144

```

```
QY 131 CACCAGACTTCCAGGAACCCGAGCTCGGCTGCTGGGAAATACTTCTGTTTGCTTC 190
Db 145 CACCAGCAATTCAGGAACCTGAGTCTGGCTGCTGGGAAATACTTCTGTTTGCTTC 204
QY 191 AACATTGTTTCTGGGTGCTGGAGCCCTGTTCTTGCCATCGGCCCTCTGGGCTGGGGT 250
Db 205 AACATTGTTTCTGGGTGCTGGAGCCCTGTTCTTGCTATCGCCCTCTGGGCTGGGGT 264
QY 251 GAGAGGGTGTCTTCCACATCTTCGCTGACCGATCTGGAGCTGGAGCCCTGACCCCTGTG 310
Db 265 GAGAGGGGCTTCTTCGACATCTCAGCGCTGACAGATCTGGAGGCTTGAACCCGCTG 324
QY 311 TGCGCTGTTGTAGTGTGGAGCGCTCATGCTCGTCTGGGCTTTCGCGCTGCATCGGG 370
Db 325 TGCGCTGTTGTAGTGTGGAGCGCTCATGCTCGTCTGGGCTTTCGCGCTGCATCGGG 384
QY 371 GCTCTCCGGAGAACACTTTCCTGCTCAAGTCTTCTCAGTGTTCCTTGCGCTCATCTTC 430
Db 385 GCCTCCGGGAGAACCTTCTGCTCAAGTCTTCTCAGTGTTCCTTGCGTCTCATCTTC 444
QY 431 TTCCTGGAGCTGGCAACGAGGATCTTGGCTTCTGATTCAGGACTGGATTCGAGACCAG 490
Db 445 TTCCTGGAGCTGGCAACGAGGATCTTGGCTTCTGATTCAGGACTGGATTCGAGACCAG 504
QY 491 CTCATTTCTTCAATTAACAACACGCTCAAGGCTATCGGATGACATTTGACCTCCAGAAC 550
Db 505 CTCATCTTTCATCAACACACGACGAGGCTTACGGGACGACATTTGACCTCCAGAAC 564
QY 551 CTCATTGACTTTCCTCAGGAATATGCTTCTGCGGAGCCCGAGGCTTAATGACTGG 610
Db 565 CTCATTGACTTTCCTCAGGAATATGCTTCTGCGGAGCCCGAGGCTTAATGACTGG 624
QY 611 AACCTCAATATCTATTCACTGACTGACTTGAACCCGAGCGGAGGCTGCGGGGTG 670
Db 625 AACCTCAATATCTATTCACTGACTGACTTGAACCCGAGCGGAGGCTGCGGGGTG 684
QY 671 CCTCTCTCTGCTGTGTCAGGACCTGTC---GATGTCTCTCAACCCAGCTGCTAT 726
Db 685 CCTCTCTCTGCTGTGTCAGGACCTGTCAGGAGGATGCTCTCAACACGATGCTGCTAC 744
QY 727 GATGTCCTG-GTCAAACTGGAGCTGG 751
Db 745 AACGTCGGAGCTCCAACTGGAGCTGG 770

RESULT 12
BX447619 888 bp mRNA linear EST 22-MAY-2003
LOCUS BX447619 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
DEFINITION clone CS0DG006YC20 5-PRIME, mRNA sequence.
ACCESSION BX447619
VERSION BX447619.1 GI:31025857
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 888)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3528.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AG006BB100P1&cluster=3528.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
```

```
Faraday Avenue Genoscope sequence ID : CS0AG006BB100P1.
FEATURES
source Location/Qualifiers
1..888
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG006YC20"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 148 a 238 c 263 g 199 t 40 others
ORIGIN
Query Match 24.1%; Score 611.6; DB 13; Length 888;
Best Local Similarity 86.7%; Pred. No. 3.3e-72;
Matches 684; Conservative 35; Mismatches 60; Indels 7; Gaps 3;
QY 12 GCCGAGCGCGCGGCTAGGCGCGGCTTAGCCCGGCGGCTTAGCCAGGCGGCGGCTGAGGCGCG 71
Db 125 GCYKCAKCGCGGCTAGGCGCGGCTTAGCCCGGCGGCTTAGCYCAGKCGGCGYCGCKGGCGCTG 184
QY 72 ATCCCGGC-CCCGGCTCCGGTTCGCCGCGGCGGCGGCTGTCTACCAATGCCGGGCAAG 130
Db 185 GGCCTGGYTCGGGCTCCGGTTTCGGGCGGCGGCTKGCCTCAMCATKCTCGGCAAG 244
QY 131 CACGACACTCCAGGAACCCGAGGCTCGGCTGCGGGGAAATACTTCTGTTTGGCTTC 190
Db 245 CACCAACATTTTAYGAAATGAGGTYGGCTGCTGCGGAAATAMTTTGTGTTGGYTTT 304
QY 191 AACATTGTTTCTGGGCTGCTGGGAGCCCTGTTCTCGCCATCGGCCCTCTGGGCTTGGGGT 250
Db 305 AAHAATTTTTCCTGGAGCTCTTCTTCTGGGCTTCTGGGCTCTGAGGCTCTGAGGCTGGGGT 364
QY 251 GAGAGGGGTGTTCTCTCAACATCTCTGCGTGAACCGATCTGGAGGCTTCGACCTGTG 310
Db 365 GAGAGGGGCTTCTCTYGAAMATCTYACGCTGACAGATCTGGGAGGCTTTGACCCGCTG 424
QY 311 TCGCTGTTGTAGTGTGGAGGCGCTCATGCTCGGCTGCTGGGCTTTGCGGCTGCATCGG 370
Db 425 TGGCTGTTGTGTTAGTGTGGAGGCGCTCATGCTCGGCTGCTGGGCTTTGCTGCTCATTTGG 484
QY 371 GCTCTCCGGGAGAACACTTTCCTGCTCAAGTCTTCTCAGTGTTCCTTGGGCTCATCTTC 430
Db 485 GCCTCCGGGAGAACACTTCTCTGCTCAAGTCTTCTCAGTGTTCCTTGGCTCATCTTC 544
QY 431 TTCCTGGAGCTGGCAACAGGATCTTGGGCTTCTGATTCAGGACTGGATTCGAGACCAG 490
Db 545 TTCCTGGAGCTGGCAACAGGATCTTGGGCTTCTGATTCAGGACTGGATTCGAGACCAG 604
QY 491 CTCATTTCTTCAATTAACAACCTCAAGGCTATCGGATGACATTTGACCTCCAGAAC 550
Db 605 CTCACCTCTTTCATCAACACACGCTCAAGGCTTACCGGAGCGACATTTGACCTCCAGAAC 654
QY 551 CTCATTGACTTTCCTCAGGAATATTTGCTCTGCGGAGCCCGAGGCGCTTAATGACTGG 610
Db 665 CTCATTGACTTTCCTCAGGAATATTTGCTCTGCGGAGGCC--GAGGCCCAATGACTGG 722
QY 611 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCGGAGGCGCTGCGGGGTG 670
Db 723 AACCTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCGGAGGCGCTGCGGGGTG 782
QY 671 CCCTCTCTCTGCTGTGTCAGGACCCCTGCG----GATGCTCTCAACACCGAGTGTGGCTAT 726
Db 783 CCCTCTCTCTGCTGTGTCAGGACCCCTGCGGAGGATGTCTCAACACCGAGTGTGGCTAC 842
QY 727 GATGTCCTGGCTCAAACTGGAGCTGGAGCGAGGCGCTCCATACCA 772
Db 843 GACGTCCTGGCTCAAACTGGAGCTGGAGCGAGGCGCTTCATCCACA 888
```

```

RESULT 13
BG770931
LOCUS
DEFINITION
602719148F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4839779 5',
mRNA sequence.
ACCESSION
BG770931
VERSION
BG770931.1 GI:14081584
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 781)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM169 row: o column: 12
High quality sequence stop: 697.
Location/Qualifiers
1..781
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4839779"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_60"
/Note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfII (sgccgctcgccc); Site_2: SfII (ggccattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGGCGGCGCATG-dT(30)BN-3'
(Where B = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT
127 a 239 c 239 g 176 t
ORIGIN
Query Match 23.9%; Score 606.4; DB 12; Length 781;
Best Local Similarity 89.4%; Pred. No. 1.8e-71;
Matches 676; Conservative 0; Mismatches 76; Indels 4; Gaps 2;
QY 12 GCCGAGCGCGGCTAGCGCCCGGGCGCTCTAGCCAGGCGCGCGTGGAGGCGG 71
DB 22 GCCGAGCGCGCGGCTAGCGCCCGGGCGGCTCTAGCCAGGCGCGCGGCGGCTGG 81
QY 72 ATCCCGGGCCCGGCTCCGGTTCCCGGGCGCGGGCGGCTGTCTACCATGCGCGGCAAGC 131
DB 82 GCCTGGCTCCCGGCTCCGGTTCCCGGGCGCGGGTGGCGCTCTACCATGCGCGCAAGC 141
QY 132 ACCAGCATCTCAGGAACCCGAGTCTGGTGTCTGGGGAATACCTCTCTTTTGGCTTCA 191
DB 142 ACCAGCATTTCCAGGAACCTGAGTCTGGTGTCTGGGGAATACCTCTCTTTTGGCTTCA 201
QY 192 ACATTGTTTCTGGGTGTGGGAGCCCTGTTCTCTGGCCATGGCCTCTGGGCTGGGGTG 251
DB 202 ACATTGTTCTTGGGTGTGGGAGCCCTGTTCTCTGGCTATCGGCTCTGGGCTGGGGTG 261

```

```

QY 252 AGAAGGTTGTTCTCTCCACATCTCTCGCTGACCGATCTGGAGGSCCTCGACCCCTGTGT 311
DB 262 AGAAGGCGGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGAGGCGCTTGACCCCGTGT 321
QY 312 GCGTGTGTTAGTGGTTGGAGCGTCATCTCGTGTCTGGGCTTTTGGCGCTGATCGGGG 371
DB 322 GCGTGTGTTAGTGGTTGGAGCGTCATCTCGTGTCTGGGCTTTTGGCTGATCGGGG 381
QY 372 CTCTCGGGAGAACACTTCTCTGCTCAAGTTTTTCTCAGTGTCTTGGCTCTCATCTTCT 431
DB 382 CCCTCGGGAGAACACTTCTCTGCTCAAGTTTTTCTCAGTGTCTTGGCTCTCATCTTCT 441
QY 432 TCCTCGAGCTGCAACAGGATCTTGGCCCTCGTATTCAAGGACTCGGATTCGAGACCAGC 491
DB 442 TCCTCGAGCTGCAACAGGATCTTGGCCCTCGTATTCAAGGACTCGGATTCGAGACCAGC 501
QY 492 TCAATTTCTTCAATTAACACACGTCAGGCGCTATCGGATGACATTTGACCTCCAGAAC 551
DB 502 TCAACCTCTTCAATTAACACACGTCAGGCGCTATCGGATGACATTTGACCTCCAGAAC 561
QY 552 TCATTGACTTTGCTCAGGAATATTGGTCTTGTCTGGGAGCGCGGAGGCGCTATGACTGGA 611
DB 562 TCATTGACTTTGCTCAGGAATATTGGTCTTGTCTGGGAGCGCGGAGGCGCTATGACTGGA 621
QY 612 ACCTCAATATCTATTTCAGTCACTGACTTGAACCCGAGCGGAGCGCTCGGGGCTGC 671
DB 622 ACCTCAATATCTATTTCAGTCACTGACTTGAACCCGAGCGGAGCGCTCGGGGCTGC 681
QY 672 CCTTCTCTGCTGTGTCTGAGGACCTCGG---ATGTCCTCAACACCCAGTGTGGCTATGA 728
DB 682 CCTTCTCTGCTGTGTCTGAGGACCTCGGAGGAGTGTCTCAACCCAGTGTGGCTATGA 741
QY 729 TGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGAGGCTC 764
DB 742 CGTCCGG-TCAACTGGAGCTGGAGCAGGAGGCTC 776
RESULT 14
BG770931
LOCUS
DEFINITION
603182305F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246304 5',
mRNA sequence.
ACCESSION
BG770931
VERSION
BG770931.1 GI:16198834
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 718)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11621 row: j column: 01
High quality sequence stop: 712.
Location/Qualifiers
1..718
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246304"
/lab_host="DH10B"
/clone_lib="NIH_MGC_121"
/Note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;

```



Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH MGC Library."

BASE COUNT 115 a 222 c 217 g 164 t  
ORIGIN

Query Match 23.7%; Score 602.4; DB 12; Length 718;  
Best Local Similarity 91.3%; Pred. No. 6.3e-71;  
Matches 639; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1 CCACGGGTCCGGCCGAGCGCGGCTAGGCCCCGGGGGCTCTAGCCCGAGGGCGGCC 60  
Db 1 CAGAGCATGAGCCGAGCGCGGCTAGGCCCCGGGGGCTCTAGCCCGAGGGCGGCC 60  
QY 61 GTGAGGGCGCATCCCGGGCCCGGCTCCGGTTCCCGGGCCGGCGGCGTGTCAACCAT 120  
Db 61 CGGGCGCTGGGCTCGCTCCGGCTCCGGTTCCGGCGCGGGGTGGCGCTCAACCAT 120  
QY 121 GCGGGCAAGCACACGACATTCAGGAACCCGAGTGGCTGTCTCGGGAAATATCTTCT 180  
Db 121 GCGGGCAAGCACACGACATTCAGGAACCCGAGTGGCTGTCTCGGGAAATATCTTCT 180  
QY 181 GTTGGCTTCAACATTTCTGGGTGTGGAGCCCTGTCTGGCCATCGGCTCTG 240  
Db 181 GTTGGCTTCAACATTTCTGGGTGTGGAGCCCTGTCTGGCCATCGGCTCTG 240  
QY 241 GCGTGGGTGAGAGGGTGTCTCTCAACATCTCTGGCTGACCGATCTGGAGGCT 300  
Db 241 GCGTGGGTGAGAGGGTGTCTCTCAACATCTCTGGCTGACCGATCTGGAGGCT 300  
QY 301 CGACCTGTGTGGCTTTGTAGTGTGGAGCGCTATGTCGGTCTGGGCTTTGCCGG 360  
Db 301 TGACCCCGTGTGGCTTTGTAGTGTGGAGCGCTATGTCGGTCTGGGCTTTGCCGG 360  
QY 361 CTGCATCGGGCTCTCCGGGAGAACATTTCTCTGCTCAAGTTTCTAGTCTCTGG 420  
Db 361 CTGCATTTGGGCGCTCTCCGGGAGAACATTTCTCTGCTCAAGTTTCTCTGG 420  
QY 421 CTCTATCTTTCTGGAGCTGGCAACAGGATCTTTGGCTTTCTGATTCAGGACTGGAT 480  
Db 421 TCTCATCTTTCTGGAGCTGGCAACAGGATCTTTGGCTTTCTGATTCAGGACTGGAT 480  
QY 481 TCGAGACGAGCTCAATTTCTTCAATTAACAAACGTCAGGCTATCGGGATGACATTGA 540  
Db 481 TCGAGACGAGCTCAACCTTCTTCAATTAACAAACGTCAGGCTATCGGGATGACATTGA 540  
QY 541 CCTCAGAACCTCATTTGCTTCAGGAATATTGGTCTTCTCGGAGCCCGAGGCC 600  
Db 541 CCTCAGAACCTCATTTGCTTCAGGAATATTGGTCTTCTCGGAGCCCGAGGCC 600  
QY 601 TAATGACTGGAACTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCGAGCG 660  
Db 601 CAATGACTGGAACTCAATATCTATTTCAACTGCACTGACTTGAACCCGAGCGAGCG 660  
QY 661 CTGCGGGTGGCTTCTCTGCTGTGTGAGGAACCTCTGG 700  
Db 661 CTGCGGGTGGCTTCTCTGCTGTGTGAGGAACCTCTGG 700

RESULT 15  
BG478644  
LOCUS  
DEFINITION 60252532F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:4643959 5',  
mRNA sequence.  
ACCESSION BG478644  
VERSION BG478644.1 GI:13410923  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 902)  
REFERENCE NIH-MGC http://img.nci.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM1414 row: p column: 08  
High quality sequence stop: 813.

Location/Qualifiers  
1..902  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4643959"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC\_20"  
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 146 a 268 c 288 g 200 t  
ORIGIN

FEATURES  
source

1..902  
Location/Qualifiers

Query Match 22.4%; Score 568.4; DB 10; Length 902;  
Best Local Similarity 88.2%; Pred. No. 1.8e-66;  
Matches 711; Conservative 0; Mismatches 81; Indels 14; Gaps 8;  
QY 12 GCGCAGCGCGCGGCTAGGCCCCGGGGGCTCTAGCCCGAGGGCGGCCGCGGCGG 71  
Db 28 GCGCAGCGCGCGGCTAGGCCCCGGGGGCTCTAGCCCGAGGGCGGCCGCGGCGG 87  
QY 72 ATCCCGGC-CCCGGCTCCGGTTCCCGCGCGCGCGGCTCTCACCATGCCGCAAG 130  
Db 88 GGCCTGGCTCCGGTTCGGGTCGGGTCGGGTCGGGTCGGGTCGGGTCGGGTCGG 147  
QY 131 CACAGCACTTCCAGGAACCCGAGTGGCTGTCTCGGGAAATACTTCTGTTTGGCTTC 190  
Db 148 CACAGCACTTCCAGGAACCTGAGTGGCTGTCTCGGGAAATACTTCTGTTTGGCTTC 207  
QY 191 AACATTGTTTCTGGTCTGGAGCCCTGTCTCGGCATCGGCTCTGGGCTGGGCT 250  
Db 208 AACATTGTTTCTGGTCTGGAGCCCTGTCTCGGCATCGGCTCTGGGCTGGGCT 267  
QY 251 GAGAAGGGTGTCTCTCAACATCTCTCGCTGACCGAT-CTGGGAGGCTCGACCTGT 309  
Db 268 GAGAAGGGTGTCTCTCGAATCTCAGCGCTGACAGATCTCTGGAGGCTCTGAG 327  
QY 310 GTGCTGTTGTAGTGTGGAGGCTCATGTCCGTGTCTGGGCTTTGGCGGCTGCATCG 369  
Db 328 GTGCTGTTGTAGTGTGGAGGCTCATGTCCGTGTCTGGGCTTTGGCTGCTATGG 387  
QY 370 GGCTCTCGGAGAACACTTCTGCTCAAGTTTTCTCAGTGTTCCTTGGCTCATCTT 429  
Db 388 GGCTCTCGGAGAACACTTCTGCTCAAGTTTTCTCAGTGTTCCTTGGCTCATCTT 447  
QY 430 CTTCCTGGAGCTGGCAACAGGATCTTGGCTCTCGTATTCAAGGACTGGATTCAGAC 489



Db 448 CTTCTCGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGGATTGAGACCA 507  
Qy 490 GCTCAATTTCTTCATTAAACAACACGCTCAAGGCTATCGGGA-TGACATTGACCTCCAGA 548  
Db 508 GGTCAAACTCTTTCATCAACAACAAACGTCAGGGCTACCGGGACCGACATTGACCTCCAGA 567  
Qy 549 ACCTCAATTGACTTTG-CTCAGGAATATTGGTCTTGTGCGGAGCCGAGGGCCTAATGAC 607  
Db 568 ACCTCAATTGACTTTGCTCAGGAATACTGGTCTTGTGAGGCCGA-GGCCCAATGAC 626  
Qy 608 TGGAACTCAATATCTATTTCAACTGC-ACTGACTTGAACCCGAGCCGAGAGCGCTGC-G 665  
Db 627 TGGAACTCAATATCTACTTCACTGCAACTGACTTGAGACCCAGCCGAGCGCTGCAG 686  
Qy 666 GGGTGCCCTCTCCTGCTGTGTCAGGAGCCCTGCGATGTCCTCAACACCCAGTG----- 719  
Db 687 GGGTGCCCTCTCCTGCTGTGTCAGGAGCCCTGCGGAGGATGTCCTCGAGAGCCAGTG 746  
Qy 720 -TGGCTATGATGTCGGCTCAAACCTGGAGCTGGAGCAGCAGGGCTCCATACACACCAAAG 778  
Db 747 TTGGCTACCGACGTCGGTTCAACTGGAGCTGGAGCAGCAGGGCTTCATCCACAGCAAAG 806  
Qy 779 GCTGTGTGGCCAGTTTGAGAGTGG 804  
Db 807 GCTGCGTGGGCCAGTTGGAGAGTGG 832

Search completed: November 21, 2003, 18:57:42  
Job time : 5188 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 15:34:38 ; Search time 654 Seconds  
(without alignments)  
10475.801 Million cell updates/sec

Title: US-09-972-970-2

Perfect score: 2538

Sequence: 1 ccacgcgtccgcgcagcc.....aaaaaaaaaaaaaaaaaaaaa 2538

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2538	100.0	2538	24 AAD21883
2	1019	40.1	2091	21 AAA96482
3	948.4	37.4	1644	24 ABL90838
4	790.4	31.1	3185	24 AAI72633
5	787.2	31.0	1126	22 AAC90014
6	773.4	30.5	1023	24 ABZ11816
7	511	20.1	1567	23 ACC44092
8	449.2	17.7	1735	22 AAH98169

9	441.6	17.4	1655	21 AAC98195	Human colon cancer
10	418.6	16.5	1174	22 AAC90015	Clone HB8EJ16 codi
11	402.2	15.8	1178	22 AAC90020	Clone HB8EJ16 codi
12	370	14.6	1988	24 ABK35735	cDNA sequence #126
13	357.4	14.1	2672	21 ABK35735	Human secreted pro
14	354.8	14.0	2465	25 ABX70953	Novel human cDNA 8
15	353.4	13.9	813	22 AAF90629	Human TANGO 339 CD
16	353.4	13.9	2715	22 AAF90628	Human TANGO 339 CD
17	351.8	13.9	813	22 AAF90656	Human TANGO 339 K1
18	351.8	13.9	813	22 AAF90657	Human TANGO 339 F2
19	351.8	13.9	813	22 AAF90658	Human TANGO 339 A4
20	351.8	13.9	813	22 AAF90659	Human TANGO 339 D6
21	351.8	13.9	1388	24 AAD29623	Human tetraepan pr
22	351.8	13.9	2623	23 AAS92016	Human tetraepanin-
23	343.8	13.5	1428	21 AAA93623	DNA encoding novel
24	333.6	13.1	1110	20 AAZ20854	Polynucleotide seq
25	333.6	13.1	1110	22 AAS59291	Human cDNA encodin
26	333.6	13.1	1110	24 ABA90960	Human polynucleoti
27	321	12.6	1667	21 AAZ65350	Human secreted pro
28	295.8	11.7	864	24 AAI72287	NET-4 antisense mo
29	294.2	11.6	368	21 AAZ93364	Sequence encoding
30	294.2	11.6	368	24 AAL41055	cDNA of Human F-bo
31	254.4	10.0	816	21 AA93675	Human tetraepanin-
32	244.6	9.6	394	22 AAF67433	Novel human polynu
33	241.8	9.5	601	23 AAS80667	DNA encoding novel
34	240	9.5	839	22 AAH34327	Human colon cancer
35	234	9.2	1932	21 AAZ98047	Human secreted pro
36	234	9.2	1932	22 AAD11660	Human secreted pro
37	234	9.2	1932	24 ABK69756	Human secreted pro
38	234	9.2	2384	24 ABZ11836	Human polynucleoti
39	233.2	9.2	852	22 ABA44553	Human cDNA encodin
40	233	9.2	389	22 AAF67421	Novel human polynu
41	225	8.9	2638	24 ABL90058	Human polynucleoti
42	174.2	6.9	353	21 AAZ42460	Human 5' EST isola
43	166.6	6.6	1120	23 ABL16799	Human 5' EST isola
44	158.2	6.2	810	23 ABL14835	Drosophila melanog
45	156	6.1	1214	23 AAS80668	DNA encoding novel

#### ALIGNMENTS

##### RESULT 1

AAZ21883

ID AAD21883 standard; cDNA; 2538 BP.

XX AAD21883;

XX AC

XX 12-FEB-2002 (first entry)

XX DT

XX DE Human TM4SF receptor-encoding gene 1 cDNA clone HOF055, SEQ ID NO:2.

XX KW Human; 4-transmembrane superfamily; TM4SF receptor; breast cancer;

XX KW ovarian cancer; immune disorder; Addison's disease; wound healing;

XX KW gene therapy; autoimmune haemolytic anaemia; autoimmune thyroiditis;

XX KW diabetes mellitus; allergy; Crohn's disease; multiple sclerosis;

XX KW rheumatoid arthritis; ulcerative colitis; cardiovascular disorder;

XX KW myocardial ischemia; neurological disease; antifungal; antiviral;

XX KW antibacterial; cerebral anoxia; epilepsy; infectious disease;

XX KW antiparasitic; cancer; ss.

XX OS Homo sapiens.

XX FH Key

XX CDS Location/Qualifiers

FT 119..820

FT /\*tag= a

FT /product= "Human TM4SF receptor protein"

XX WO200177173-A1.

XX PD 18-OCT-2001.

XX 05-APR-2001; 2001WO-US11130.

XX 10-APR-2000; 2000US-195336P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Shi Y, Ruben SM;  
XX WPI; 2002-017447/02.  
DR P-PSDB; AAE13218.  
XX Novel isolated protein, a member of 4-transmembrane superfamily of  
PT receptor polypeptides, useful in the prevention, treatment and  
PT diagnosis of cancer, immune disorders, cardiovascular disorders and  
PT neurological diseases -  
XX Claim 1; Page 265-266; 271pp; English.  
XX The invention relates to human 4-transmembrane superfamily (TM4SF)  
CC receptor polypeptides and polynucleotides. Sequences of the invention  
CC are useful for preventing, treating, ameliorating or diagnosing a  
CC pathological condition or a susceptibility to a pathological condition.  
CC TM4SF polypeptides are useful for screening molecules which modify  
CC their activity. TM4SF nucleic acids, protein, antibodies, agonists and  
CC antagonists are useful in the diagnosis, treatment and prevention of  
CC cancer, particularly breast and ovarian cancer, and other cancers of  
CC the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,  
CC liver, lung, or urogenital; immune disorders such as Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemia; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections. TM4SF polynucleotides are  
CC also useful in gene therapy. The present sequence is human TM4SF  
CC receptor protein encoding cDNA.  
XX  
SQ Sequence 2538 BP; 528 A; 669 C; 747 G; 594 T; 0 other;  
  
Query Match 100.0%; Score 2538; DB 24; Length 2538;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 CCACGGCTCCGGCCGAGCGCCGGCTAGGCCCGGGGGCTCTAGCCAGGGCGGCC 60  
DB 1 CCACGGCTCCGGCCGAGCGCCGGCTAGGCCCGGGGGCTCTAGCCAGGGCGGCC 60  
  
QY 61 GTGGAGGGCGATCCCGGCCCGGGCTCCGGTCCCGGGCGGGCGGCTCTCACCAT 120  
DB 61 GTGGAGGGCGATCCCGGCCCGGGCTCCGGTCCCGGGCGGGCGGCTCTCACCAT 120  
  
QY 121 GCGGGCAAGCACACGACCTCCAGAACCCGAGGTCGGCTCTGCGGGAAATCTTCT 180  
DB 121 GCGGGCAAGCACACGACCTCCAGAACCCGAGGTCGGCTCTGCGGGAAATCTTCT 180  
  
QY 181 GTTGGCTTCAACATTTTCTGGGTCTGGAGCCCTGTCTCGGCCATCGGCTCTG 240  
DB 181 GTTGGCTTCAACATTTTCTGGGTCTGGAGCCCTGTCTCGGCCATCGGCTCTG 240  
  
QY 241 GGCCTGGGTGAGAGGGTGTCTCTCAACATCTCTCGCTGACCGCATCTGGAGGCT 300  
DB 241 GGCCTGGGTGAGAGGGTGTCTCTCAACATCTCTCGCTGACCGCATCTGGAGGCT 300  
  
QY 301 CGACCTGTGTGGCTCTTGTAGTGTGGAGCGCTATGTCGGCTTTGGCGG 360  
DB 301 CGACCTGTGTGGCTCTTGTAGTGTGGAGCGCTATGTCGGCTTTGGCGG 360  
  
QY 361 CTGCATCGGGCTCTCCGGGAGAACATTTCTCTGTCTCAAGTTTCTCAGTGTCTTGG 420  
DB 361 CTGCATCGGGCTCTCCGGGAGAACATTTCTCTGTCTCAAGTTTCTCAGTGTCTTGG 420  
  
QY 421 CCTCATCTTCTTCTGGAGCTGGCAACAGGGATCTTGGCTTCTCGTATTCAGGACTGGAT 480  
DB 421 CCTCATCTTCTTCTGGAGCTGGCAACAGGGATCTTGGCTTCTCGTATTCAGGACTGGAT 480

QY 481 TCAGACCAAGCTCAATTTCTTCAATTAAACAACAAGTCAAGGCTATCGGATGACATGA 540  
DB 481 TCAGACCAAGCTCAATTTCTTCAATTAAACAACAAGTCAAGGCTATCGGATGACATGA 540  
  
QY 541 CCTCCAGAACCTCAATTGCTCTCAGGAATATTGGTCTTGTGGGAGCCCGAGGGCC 600  
DB 541 CCTCCAGAACCTCAATTGCTCTCAGGAATATTGGTCTTGTGGGAGCCCGAGGGCC 600  
  
QY 601 TAATGACTGGAACCTCAATATCTATTTCAACTGCACCTGACTTGAACCCGAGCGAGAGCG 660  
DB 601 TAATGACTGGAACCTCAATATCTATTTCAACTGCACCTGACTTGAACCCGAGCGAGAGCG 660  
  
QY 661 CTGCGGGTGCCTTCTCTCTGCTGTCTCAGGAGCCCTGCGATGTCTCAACACCCAGTGT 720  
DB 661 CTGCGGGTGCCTTCTCTCTGCTGTCTCAGGAGCCCTGCGATGTCTCTCAACACCCAGTGT 720  
  
QY 721 GGTATGATGTCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACCAAGGC 780  
DB 721 GGTATGATGTCGGCTCAAACTGGAGCTGGAGCAGCAGGGCTCCATACACCAAGGC 780  
  
QY 781 TGTGTGGGCGAGTTTTCAGAAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGCTCTT 840  
DB 781 TGTGTGGGCGAGTTTTCAGAAAGTGGCTGCAGGACAACCTGATCGTGGTGGCTGGGCTCTT 840  
  
QY 841 GTGGGATCGCTCTCTCCAGATCTTTGGTATCTGCTGGCCCGAGAACCTTGTGAGTGAC 900  
DB 841 GTGGGATCGCTCTCTCCAGATCTTTGGTATCTGCTGGCCCGAGAACCTTGTGAGTGAC 900  
  
QY 901 ATCAAGGCGAGTGAAGCCCAACTGGATCAAAACATGATGATGGCTACAACTACTCAATAA 960  
DB 901 ATCAAGGCGAGTGAAGCCCAACTGGATCAAAACATGATGATGGCTACAACTACTCAATAA 960  
  
QY 961 ACAAAACCTTGAACCACTGGCTTACGCCCACTCTCAGAGGTTCCATGGGCGCAGG 1020  
DB 961 ACAAAACCTTGAACCACTGGCTTACGCCCACTCTCAGAGGTTCCATGGGCGCAGG 1020  
  
QY 1021 GCTCAGCCGCTGCTGCTGGGCGCCAGCCAGACCCAGCCCTGCGCAACATGTTTTC 1080  
DB 1021 GCTCAGCCGCTGCTGCTGGGCGCCAGCCAGACCCAGCCCTGCGCAACATGTTTTC 1080  
  
QY 1081 TTGGCTCTGGTGTAGTACATGATGAGCAACCTTTTAAACTTGGCATATTTATGATGATAA 1140  
DB 1081 TTGGCTCTGGTGTAGTACATGATGAGCAACCTTTTAAACTTGGCATATTTATGATGATAA 1140  
  
QY 1141 GTCCAGATCCCGAGCATCTTGTGAAGATGGCCATCCGGCCACAGCGGCTCTTCTATGGC 1200  
DB 1141 GTCCAGATCCCGAGCATCTTGTGAAGATGGCCATCCGGCCACAGCGGCTCTTCTATGGC 1200  
  
QY 1201 TTGCTCTCTGGGATGTGGCTTCTGTTCTCTGAGGAGCCACCCCTCACCCGTGCTG 1260  
DB 1201 TTGCTCTCTGGGATGTGGCTTCTGTTCTCTGAGGAGCCACCCCTCACCCGTGCTG 1260  
  
QY 1261 CTGCTCTGACCTGGAGGCTGGAGCTGGCCCTCTCCACTCTGCAAGTTTTTCCCTGC 1320  
DB 1261 CTGCTCTGACCTGGAGGCTGGAGCTGGCCCTCTCTCCACTCTGCAAGTTTTTCCCTGC 1320  
  
QY 1321 AAATGCTCAAGGCTCTGTGGGCAAGCCCGGATCGAAGCCCTGAGCGTGAAGAATTTG 1380  
DB 1321 AAATGCTCAAGGCTCTGTGGGCAAGCCCGGATCGAAGCCCTGAGCGTGAAGAATTTG 1380  
  
QY 1381 GAGAGCTGGAGCTGCTCCCAAGAGGCGACAGCTTGGGAAGGGTCTGGGCTCTCTGGGG 1440  
DB 1381 GAGAGCTGGAGCTGCTCCCAAGAGGCGACAGCTTGGGAAGGGTCTGGGCTCTCTGGGG 1440  
  
QY 1441 CCAAGATGGGTGCCACCGTGGCCAGAGTGGCGGGTGGGATGAGATCAGGAAG 1500  
DB 1441 CCAAGATGGGTGCCACCGTGGCCAGAGTGGCGGGTGGGATGAGATCAGGAAG 1500  
  
QY 1501 GTTTTGGGCGAGCAGCTAGCTGGAAGCTGAGCTTGTCACCCATGGGGATGGGAGAGCCC 1560  
DB 1501 GTTTTGGGCGAGCAGCTAGCTGGAAGCTGAGCTTGTCACCCATGGGGATGGGAGAGCCC 1560

1561 TGTTTGGAGCGCTGATGGTATGAGTCTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTCC 1620  
1561 TGTTTGGAGCGCGCTGATGGTATGAGTCTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTCC 1620  
1621 CAGTGGCTGTAGATGCTGCTGACACATAGAGCGCGCGCCACACAGCATATCCTGG 1680  
1621 CAGTGGCTGTAGATGCTGCTGACACATAGAGCGCGCGCCACACAGCATATCCTGG 1680  
1681 GTCTGCTTTCTTAGGACAAACCCACCCAGTACAGCCCTGTGCTGTGTGTCACCCCTG 1740  
1681 GTCTGCTTTCTTAGGACAAACCCACCCAGTACAGCCCTGTGCTGTGTGTCACCCCTG 1740  
1741 CTTACTAGTCTTTGGTTTCATGGAAATTTACAAGCTTCTAAAGGACGAGTGGCTCAG 1800  
1741 CTTACTAGTCTTTGGTTTCATGGAAATTTACAAGCTTCTAAAGGACGAGTGGCTCAG 1800  
1801 ATTGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGGAGTATTG 1860  
1801 ATTGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGGAGTATTG 1860  
1861 TGAATCAAGAGAGAGGTTTGGCTAGTGCCTAGTCTTTTAACTTAGATGCCCTCAGGGCC 1920  
1861 TGAATCAAGAGAGAGGTTTGGCTAGTGCCTAGTCTTTTAACTTAGATGCCCTCAGGGCC 1920  
1921 GGGTGGGTTATAAAATAAAGTAGGCTTTGAGCTGTGAGGCTTTGGACTTTAATTTT 1980  
1921 GGGTGGGTTATAAAATAAAGTAGGCTTTGAGCTGTGAGGCTTTGGACTTTAATTTT 1980  
1981 TCCGACTATCTCGAGATGGGACATAGAGACATGCTTTGTGCTGAGAAATCTTGC 2040  
1981 TCCGACTATCTCGAGATGGGACATAGAGACATGCTTTGTGCTGAGAAATCTTGC 2040  
2041 ATGATTGAGTCTGAGTCTGAGTCTGAGGCACTGCGCTTGAAGTACATCAAGGGGTGTGGG 2100  
2041 ATGATTGAGTCTGAGTCTGAGTCTGAGGCACTGCGCTTGAAGTACATCAAGGGGTGTGGG 2100  
2101 ACTGTGGCAAAACACAGATTTCCACCTGAAATTTGGTGGCTGCTTCCGTTGGGCTAAT 2160  
2101 ACTGTGGCAAAACACAGATTTCCACCTGAAATTTGGTGGCTGCTTCCGTTGGGCTAAT 2160  
2161 GGCTGTACAGCAGATGATGATGATGCTGCTAATGGGAGAGTCTGGAGAACCAAGA 2220  
2161 GGCTGTACAGCAGATGATGATGATGCTGCTAATGGGAGAGTCTGGAGAACCAAGA 2220  
2221 AGCTAGATTTTTCATGTGAATATCCCGAGTTTAAAGTTTGTGAGTCTTAATGAGAAAA 2280  
2221 AGCTAGATTTTTCATGTGAATATCCCGAGTTTAAAGTTTGTGAGTCTTAATGAGAAAA 2280  
2281 CCTCTTAAACCTGTATGATCAAAAGGTGTGGGGCCCATCTTTGACACCTTCCGCCACCAT 2340  
2281 CCTCTTAAACCTGTATGATCAAAAGGTGTGGGGCCCATCTTTGACACCTTCCGCCACCAT 2340  
2341 GGTCCCTCAGGACAGTCCCATGCGGAGCCCTGTGAGTCAACGGTTCCAGGAATGCTGG 2400  
2341 GGTCCCTCAGGACAGTCCCATGCGGAGCCCTGTGAGTCAACGGTTCCAGGAATGCTGG 2400  
2401 GAACTGTGAGGTTGGGGGTTGGGCGAGCAGCCATCGTGGCTCCAGTGTGGGCT 2460  
2401 GAACTGTGAGGTTGGGGGTTGGGCGAGCAGCCATCGTGGCTCCAGTGTGGGCT 2460  
2461 GGGTGGGCTATGGTGGGCTTACATGCTACAATAAATGGGGCTCATGATAAAAAAAA 2520  
2461 GGGTGGGCTATGGTGGGCTTACATGCTACAATAAATGGGGCTCATGATAAAAAAAA 2520  
2521 AAAAAAAAAAAAAAAAAAAAA 2538  
2521 AAAAAAAAAAAAAAAAAAAAA 2538

RESULT 2  
AAA96482  
ID AAA96482 standard; cDNA; 2091 BP.  
XX

AC AAA96482;  
XX 08-FEB-2001 (first entry)  
DT  
XX cDNA encoding a human transmembrane protein.  
DE Human; transmembrane protein; cell proliferation disorder; myeloma;  
KW reproductive disorder; smooth muscle disorder; neurological disorder;  
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;  
KW Alzheimer's disease; Tourette's disorder; ss.  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
FH CDS 113..1102  
FT /\*tag= a  
XX  
XX WO200056891-A2.  
XX  
XX 28-SEP-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07817.  
XX  
XX 22-MAR-1999; 99US-0125537.  
PR 16-JUN-1999; 99US-0139565.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O, Baughn MR;  
PI Lu DAM, Azimzai Y, Yang J;  
XX  
XX WPI: 2000-579485/54.  
DR P-PSDB; AAB18969.  
XX  
XX New human transmembrane proteins are used to treat a disease or  
PT condition associated with decreased expression of functional HTMP e.g.  
PT Tourette's disorder, angina and leukaemia -  
XX  
XX Claim 4; Page 113-114; 130pp; English.  
XX  
XX The present sequence encodes a human transmembrane proteins (HTMP).  
CC Agonists and antagonists of the protein are used to treat a disease  
CC or condition associated with overexpression of the protein. Diseases  
CC and conditions which can be treated include cell proliferative,  
CC immunological, reproductive, smooth muscle and neurological disorders  
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency  
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,  
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The  
CC polynucleotides may be used to detect and quantify gene expression in  
CC biopsied tissues where protein expression may be correlated with disease  
CC e.g. to determine absence, presence or excess expression of HTMP or to  
CC monitor regulation of HTMP expression during therapeutic intervention.  
XX  
SQ Sequence 2091 BP; 392 A; 602 C; 637 G; 460 T; 0 other;  
  
Query Match 40.1%; Score 1019; DB 21; Length 2091;  
Best Local Similarity 76.6%; Pred. No. 3.3e-236;  
Matches 1544; Conservative 0; Mismatches 365; Indels 106; Gaps 20;  
  
QY 12 GCCCGAGCGCGCGCTAGGCGCGCGCGCTCTAGCCAGCGCGCGCGCTGAGGCGCG 71  
Db 5 GCCCGAGCGCGCGCGCTAGGCGCGCGCGCTCTAGCCAGCGCGCGCGCGCGCTG 64  
QY 72 ATCCCGGC-CCCGGCTCCGGTTCCCGGCGCGCGCGCGCTGCTCACCATGCCGGCAAG 130  
Db 65 GGCCTGGCTCCCGGCTCCGGTTCCCGGCGCGCGCGCGCTGCTCACCATGCCGGCAAG 124  
QY 131 CACCAGACCTTCCAGGAACCCAGGTCGGTGTGCTGCGGGAATACTTCTGTTGGCTTC 190  
Db 125 CACCAGACCTTCCAGGAACCCAGGTCGGTGTGCTGCGGGAATACTTCTGTTGGCTTC 184  
QY 191 AACATTGTTTCTGGGTGCTGGGAGCCCTGTTCTCGGCCATCGGCGCTCTGGGCTTGGGGT 250

185 AACATTGCTCTTGGGTGCTGGAGCCCTGTTCTGGCTATCGGCTCTGGGCTTGGGT 244  
251 GAGAGGGGTGTTCTCTCAACATCTCTGCTGACCGATCTGGAGGCTTCGACCTGTG 310  
245 GAGAGGGGGTCTCTCGAACATCTCAGCGCTGACAGATCTGGAGGCTTCGACCGGTG 304  
311 TGGCTGTTTGTAGTGTGGAGCGCTCATGTCGCTGCTGGGCTTGGCGGCTGCATCGG 370  
305 TGGCTGTTTGTAGTGTGGAGCGCTCATGTCGCTGCTGGGCTTGGCTGGCTGCATGG 364  
371 GCTCTCGGAGAACACTTCTCTGCTCAAGTTTCTCAGTGTTCCTTGGCTCATCTTC 430  
365 GCTCTCGGAGAACACTTCTCTGCTCAAGTTTCTCTGCTGTCTCTGCTCATCTTC 424  
431 TTCTGGAGCTGGCAACAGGATCTTGGCTTCTGATTCAGAGACTGATTCGAGACCG 490  
425 TTCTGGAGCTGGCAACAGGATCTTGGCTTCTGATTCAGAGACTGATTCGAGACCG 484  
491 CTCAATTTCTTCAATTAACAAACAGCTCAAGGCTTATCGGGATGACATTGACCTCCAGAAC 550  
485 CTCACCTTCTCATCAACAAACAGCTCAAGGCTTACCGGAGCAGATTGACCTCCAGAAC 544  
551 CTCATTGACTTGTCTCAGGAATATGCTTCTGCGAGCCCGAGGCTTAATGACTGG 610  
545 CTCATTGACTTGTCTCAGGAATATGCTTCTGCGAGCCCGAGGCTTAATGACTGG 604  
611 AACCTCAATATCTTCAACTGCACTGATTTGAACCCGAGCCGAGGCTTGGGGGTG 670  
605 AACCTCAATATCTTCAACTGCACTGATTTGAACCCGAGCCGAGGCTTGGGGGTG 664  
671 CCCTTCTCTGCTGTGTAGGAGCCCTGCG --- GATGCTCTCAACACCCAGTGGCTAT 726  
665 CCCTTCTCTGCTGTGTAGGAGCCCTGCGAGGATGCTCTCAACCCAGTGGCTAT 724  
727 GATGCTCGGCTCAAACTGAGCTGAGCAGAGGCTCATACACCAAAAGGCTGTGTG 786  
725 GAGCTCCGCTCAAACTGAGCTGAGCAGAGGCTCATCAACCAAAAGGCTGCGTG 784  
787 GSCCAGTTTGAAGTGGCTGAGCAGACCTGATGCTGGTGGCTTGGGTCTTGTGGC 846  
785 GSCCAGTTTGAAGTGGCTGAGCAGACCTGATGCTGGTGGCTTGGGTCTTGTGGC 844  
847 ATGCTCTCTCCAGATCTTGGTATCTGCTGGCCAGAACCTTGTGAGTGACATCAAG 906  
845 ATGCTCTCTCCAGATCTTGGTATCTGCTGGCCAGAACCTTGTGAGTGACATCAAG 904  
907 CGAGTGAAGCCCACTGGATCAAAATGATGATGCTCAAACTACTCAAAATAAAACAAA 966  
905 GCAGTGAAGCCCACTGGAGCAAAATGGAATGATG-----A 939  
967 CTTGAAACCACTGGCTTACGCCACCATCTCAGAGGTTCATGGGCGGAGGCTCA 1026  
940 CTTTGAACCACTGGCTTACGCCACCATCTTCCAGAGTCTGTCCACGGGGGCTCA 999  
1027 GCGTGGCTGTCTGCTGGGGCCC - AGCCAGACCCAGCTGCTGCCAACATGTTTC - TTGG 1084  
1000 GCAGAACTCTCTGACTGGGGCCCTGCGCCGCCCCAGCCAGCATGTTTCTTTGG 1059  
1085 CTTGGGTAGTACATGATGACCAACCTTTAAACCTTGGCATATTTCAATGAAAAGTCC 1144  
1060 CTTGGGTGGTTTATACCTCGAGCAACCTTTAAACCTTGGTAGATTTTCAATGAAAAGTCC 1119  
1145 AGATCCCGAGATCTTGTGAAGATGGCATCCGCGCCAGCGGCTTCTATGCTTGG 1204  
1120 AGATCCAGATCTTGTGAAGATGACACCTGGCTAGCGGCTCTTGGTGGCAACA 1179  
1205 TCTCTGGATGTGCGCTTCTGTTCTCTGAGGACCCAGCTCAACCGGTGTC----- 1258  
1180 CTACTGGGACACTGCTCTCCAGTCAACAGGGGCCAGCTGGCCCGTCTACTCACCT 1239  
1259 ---TGCTGCTGACCTGGAGCTGGAGCTGGCTCTCTCACTCTGCAAGTTTTC 1315

1240 AAGTCCGCTGAGCCTTGTACACTAGGAGCTGGCTC---CCACCTCTCAGGGTTATTT 1297  
1316 CTTGCAAAATGTCGAAGCTGCTGTGGGCAAGCCGATCAAGCTCGAGCGTGAAGA 1375  
1298 CTTGCA---CCTCAGGCGCTGCGGCGCAATCTGGAGTGAACACACGGGACCTGAAG 1353  
1376 ATTGGGAGGCTGGAGCCTGCCCCAAAGAGGCCACAGCTGGAGGGTCTGGCCCTCT 1435  
1354 A-TGGAGAGGCTGGAGCCCGCTTTGAAGAGGGTGCAGCCTGGGAAGGG---CGGCCCTTGT 1410  
1436 GGGGCCCAAGATGGGTGCGCACCGTCCCGAGAGAGTGGCCGAGGGTGGGATGGAGATCA 1495  
1411 GGGGACTCGGCTGGAG-TAGAGTCCCGAGAGAGGCTCTGAGGGTGGGATGGGGTCA 1469  
1496 GGAAGGTTTTGGGCAAGAGCTAGCTGGAAGCCTGAGCTTGTCAACCATGGGGATGGGAG 1555  
1470 GGACAAATTTTGCAAAAGAAAGTAGCTGGAAGCC-----ATGGGACTGGCG 1514  
1556 AGCCCTGTTTGGAGGCGGCTGATGATGATGAGTCACTCAGCCTCTGTGGAACTCAGTCAAAAT 1615  
1515 GAGCCTGTTTGGGGGATCTGGATGGTTGACTCC-----TAGGAGTCAAGTTCAGCAT 1566  
1616 CTTCCAGTGGGCTCTAGAGTTGCTCTGACCACTAGAGGCGCGCCACACAGCATTA 1675  
1567 CTTGCGCGTGG-CTGACAGCTGCTGATGGCACTAGAGGSCAGCCAGCCACACTC 1625  
1676 CTTGGTCTGCTTTCTTAGAACACCCACAGTACAGCCTGTGCTGTGTGTCCA 1735  
1626 CTTGGTCTGCTTCTCTCCGCAACTCACTCTAGTAGGCTGTG-----1671  
1736 CCCTGTTTACTAGTCTTGTGGGTTTCATGGAAATTTACAGCTTCTAAAGGACAGTGG 1795  
1672 -CCTGCTTACTAGCCTCTGGGTTTGGAGAGTTTGGAAATTTCTCA--GAGCCAACTGG 1728  
1796 CTCAGATTGGGGAAGCCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTG-TGGAG 1854  
1729 CTCAGCTTGGGAAGCTGGCTGTGCTGCTCAGCTCCGCTCATCAGCTATGTGAAGGG 1788  
1855 TATTGTGAATCAAAAGAGAG------TTTGGCTAGTGGCCAGTCTTTTAACTTGA 1907  
1789 TGTGTGGAGTGTATCTCTGCGGCCCTCTCTGGCTGTGTCAGAGATCTCAAACTCCGA 1848  
1908 TGCCCTCAGGCGCGGTTTATAAAA---ATAAGTAGCCTTTGAGCTGTGAGGCC 1963  
1849 TGCCCTCAGGCGCGTATGTTGTAATGATGAACAGGCGCTTGTGATTGGAGCCT 1908  
1964 TTTGGGACTTTAAATTTTCCCACTATTCTCTGGAGA 1998  
1909 GCTTCACTTTGA--CTTCCCACTGTTGCTGGAGA 1941

RESULT 3  
ABL90838  
ID ABL90838 standard; cDNA; 1644 BP.  
XX ABL90838;  
AC ABL90838;  
XX  
DT 24-MAY-2002 (first entry)  
XX Human polynucleotide SEQ ID NO 1400.  
DE  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.  
OS Homo sapiens.  
XX  
XX WO200190304-A2.  
XX  
XX 29-NOV-2001.  
XX

PF 18-MAY-2001; 2001WO-US16450.  
XX  
PR 19-MAY-2000; 2000US-205515P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Birse CE, Rosen CA;  
XX  
DR WPI; 2002-122018/16.  
XX P-PSDB; ABB90429.  
XX  
XX  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders -  
XX  
XX Claim 4; SEQ ID NO 1400; 2081pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to novel genes (ABL9449-ABL90853) and proteins  
CC (ABB90440-ABB90444) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and antagonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune,  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 1644 BP; 297 A; 479 C; 513 G; 346 T; 9 other;

Query Match 37.4%; Score 948.4; DB 24; Length 1644;  
Best Local Similarity 82.1%; Pred. No. 3.6e-219;  
Matches 1244; Conservative 6; Mismatches 213; Indels 52; Gaps 12;

QY 12 GCCGAGCCGCGGCTAGCCGCGGCGCTCTAGCCAGGCGCGCGCTGAGGCGCG 71  
DB 114 GCCGAGCCGCGGCTAGCCGCGGCGCTCTAGCCAGGCGCGCGCGCGCGCTG 173  
QY 72 ATCCCGGC-CCCGGCTCCGCTTCCCGGCGCGCGCGCGCTGCTCACCATGCCGCGCAAG 130  
DB 174 GGCCTGGCTCCCGGCTCCGCTTCCCGGCGCGCGCGCGCTCACCATGCCGCGCAAG 233  
QY 131 CACCAGCACTTCCAGGACCCGAGGTCGGCTGCTGCGGGAATACTTCTGTTGGCTTC 190  
DB 234 CACCAGCACTTCCAGGACCCGAGGTCGGCTGCTGCGGGAATACTTCTGTTGGCTTC 293  
QY 191 AACATTGTTTCTGCGTGTGGAGCCCTGTTCTCGCCATCGGCTCTGGGCGCTGGGT 250  
DB 294 AACATTGTTTCTGCGTGTGGAGCCCTGTTCTCGGCTATCGGCTCTGGGCGCTGGGT 353  
QY 251 GAGAGGGGTGTTCTTCCCAATCTCTGCGTGAACCGATCTGGGAGCCCTCGACCTGTG 310  
DB 354 GAGAGGGGTGTTCTTCCCAATCTCTGCGTGAACCGATCTGGGAGCCCTCGAMCCGTG 413  
QY 311 TGGCTGTTTGTAGTGGTGGAGCGTCATGTCGGTCTGGGCTTTCGGGCTGCATCGGG 370  
DB 414 TGGCTGTTTGTAGTGGAGCGTCATGTCGGTCTGGGCTTTCGGGCTGCATTGGG 473  
QY 371 GCTCTCCGGGAGAACACTTCTGCTCAAGTTTCTCAGTGTTCCTTGGCTCTCATCTTC 430  
DB 474 GCCCTCCGGGAGAACACTTCTGCTCAAGTTTCTCAGTGTTCCTTGGCTCTCATCTTC 533  
QY 431 TTCCTGGAGCTGGCAACAGGATCTTGGCTTCGTATTCAAGGACTGGATTGAGACCAG 490

DB 534 TTCTGGAGCTGGCAACAGGGATCCTGGCCTTTGTCTTCAAGGACTGATTCGAGACCAG 593  
QY 491 CTCAATTTCTTCAATTAACAACAACGCTCAAGGCGCTATCGGGATGATGATTCAGCTCAGAAC 550  
DB 594 CTCACCTCTTCAATCAACAACAACGCTCAAGGCGCTATCGGGAGACGATTCAGCTCAGAAC 653  
QY 551 CTCATTGACTTTGCTCAGGAATATGCTCTTGGCGGAGCCCGAGGCGCTAATGACTGG 610  
DB 654 CTCAATGACTTTGCTCAGGAATATGCTCTTGGCGGAGCCCGAGGCGCTAATGACTGG 713  
QY 611 AACCTCAATATCTAATTTCAACTGCACTGACTTGAACCCGAGCCGAGAGCGCTGGGGGTG 670  
DB 714 AACCTCAATATCTAATTTCAACTGCACTGACTTGAACCCGAGCGGAGCGCTGGGGGTG 773  
QY 671 CCCTTCTCTGCTGTGTGAGGAGCCCTGC-----GATGTCTCAACACCCAGTGTGGCTAT 726  
DB 774 CCCTTCTCTGCTGTGCGTCAAGGAGCCCTGCGGAGGATGTCTCTCAACCCAGTGTGGCTAC 833  
QY 727 GATGTCCGGCTCAAACTGGAGCTGGAGCAGCGGCTCCATACACACCAAGGCTGTGTG 786  
DB 834 GAGTCCGGCTCAAACTGGAGCTGGAGCAGCGGCTTCATCCACCAAGGCTGTGTG 893  
QY 787 GGCAGGTTTGAGAGTGGCTGAGGACCACTGATCGTGTGGCTGGGCTCTTTGTGGGC 846  
DB 894 GGCAGGTTTGAGAGTGGCTGAGGACCACTGATTTGTGTGTGGCGGAGTCTTCATGGGC 953  
QY 847 ATGCTCTCTCCAGATCTTTTGGTATCTGCTGGCCAGAACCTTTGAGTGCATCAAG 906  
DB 954 ATGCGCTCTCCAGATCTTTTGGCATCTGCTGGCCAGAACCTTCGTGAGTGACATCAAG 1013  
QY 907 GCAGTGAAGCCCACTGGATCAACATGATGATGGCTTCAAACTACTCAATAAACAATA 966  
DB 1014 GCAGTGAAGCCCACTGGAGCAAAATGGAATGATG-----A 1048  
QY 967 CCTTGAAGAACCACTGGCTTACGCCCACTCTCAGAGGTTTCCATGGGCGCAGGCGCTCA 1026  
DB 1049 CTTTGAAGAACCACTGGCTTACGCCCACTCTTCCGAGGCTCTGTCACGCGGGGCGCTCA 1108  
QY 1027 GCGTGCCTGCTGCTGGGCGCC-AGCCAGACCCACCTCGCCCAACATGTTTCT-TTGG 1084  
DB 1109 GCAGAACTCTCTGACTGGGCGCCCTGGCCNGGCCACCCAGCCGACATGTTTCTTTGG 1168  
QY 1085 CTTGGGTAGTACATGAGCCCACTTTAAACTTGGCATATTTTCATGTAAGTCC 1144  
DB 1169 CTTGGGTGTTTATACCTTGAGCCAACTTTAAATAATTTGTAGATTTTCAATAAAGTCC 1228  
QY 1145 AGATCCCGAGCACTTTGTGAAGATGCGCATCCGCGCACAGCGCTCTTCTATGGCTTCG 1204  
DB 1229 AGATCCAGAGCTTCTTTGAAGATGACCACTGGGTAC-GCGGCTCTTCGGTGGCAACA 1287  
QY 1205 TCTCTGGGATGTGCGCTTCTGTTCTCTGAGGAGCCCACTCAGCCGCTGTC----- 1258  
DB 1288 CTACCTGGGACACTGCTCCCACTCAGGAGCCCGAGTGGCGCGTCTACTACCT 1347  
QY 1259 ---TGCTGCTGACCTGGAGGCTGGAGCTGGGCTCTCTCAGCTCTGCAAGTTTTC 1315  
DB 1348 AAGTGGCGCTGACCTTGTACACTAGGAGCTGGCCTC--CCACCTCTGAGGTTTATTT 1405  
QY 1316 CTTGCAAAATGCTGCAAGGCTGTGCGGCAAGCCCGGATCGAAGCTGGAGCGTGAAGA 1375  
DB 1406 CTTGCA---CTTCGAGGCGCTGCGGCGCAATCTGGAGTGAACAACGGGGAAGG 1461  
QY 1376 ATTGGGAGGCTGGAGGCTGCCCCAAAGAGGCCACAGCTCTGGGAGGCTCTGGCCCTCT 1435  
DB 1462 A-TGAGAGGCTGGAGGCTGAGGCTTTGAAGAGGCTGAGCTCTGGGAGGG--CGGCTTGTCT 1518  
QY 1436 GGGGCGCAGATGGGTGCCACCGTGGCCAGGAGAGTGGCGGAGGCTGGAGTGAATCA 1495  
DB 1519 GGGGACTCGGTGGGAG-TAGAGTTCGCCAGGARGGTCTGAGGCGGTGGATTTGGGGGKC 1577  
QY 1496 GGAAGGTTTGGGCA 1510  
DB 1578 AGGCAATTTTGGCA 1592

## RESULT 4

AAI72633 standard; cDNA; 3185 BP.

XX AAI72633;

XX 18-JUN-2002 (first entry)

XX DE Human 23228 cDNA.

XX Gene; tetraspanin; 23228; cell surface protein; transmembrane domain;  
 KW extracellular growth factor; HB-EGF; TGF-alpha; amphiregulin;  
 KW diabetes mellitus; arthritis; multiple sclerosis; encephalomyelitis;  
 KW dermatitis; Crohn's disease; and asthma; cancer; metastasis;  
 KW viral infection; cellular signalling activity; cell proliferation;  
 KW cell motility; CD81; B-Cell antigen receptor; ss.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers  
 FT CDS 168..980  
 FT /\*tag= a  
 FT /product= "23228"

XX WO200216603-A2.

XX 28-FEB-2002.

XX 21-AUG-2001; 2001WO-US41811.

XX 21-AUG-2000; 2000US-226612P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Leiby KR;

XX WPI; 2002-315466/35.

XX P-PSDB; AAB47950.

XX New tetraspanin 23228 polypeptide useful in screening assays,  
 PT predictive medicine and as a prophylactic or therapeutic agent, e.g.,  
 PT for hematopoietic and immune diseases such as diabetes or multiple  
 PT sclerosis -

XX Claim 1; Page 111-113; 123pp; English.

CC This sequence encodes tetraspanin 23228. This protein is a cell surface  
 CC protein having four transmembrane domains (TM1-TM4). TM1, TM3 and TM4  
 CC have a single polar amino acid located within them, which may interact  
 CC with each other and contribute to domain stability. The cytoplasmic N-  
 CC and C-terminal domains and the intracellular loop between TM2 and TM3  
 CC are less than 30 amino acids in length. The second extracellular domain  
 CC between TM3 and TM4 contains conserved Cys residues and may function  
 CC to bind extracellular growth factors, such as HB-EGF, TGF-alpha and  
 CC amphiregulin. The 23228 protein is useful for diagnosing and treating  
 CC 23228-mediated disorders, e.g., haematopoietic and/or immune disorders  
 CC such as diabetes mellitus, arthritis, multiple sclerosis,  
 CC encephalomyelitis, dermatitis, Crohn's disease and asthma; cell  
 CC proliferation and differentiation disorders, e.g., cancers or  
 CC metastasis; and/or viral infections. 23228 polypeptide can control  
 CC cellular signalling activity, bind to an extracellular growth factor,  
 CC for example, amphiregulin, regulate cell proliferation, bind to a cell  
 CC surface protein, to recruit intracellular kinases, to regulate cell  
 CC motility, bind to another tetraspanin such as CD81, to associate with a  
 CC B-Cell antigen receptor and the ability to modulate the association  
 CC with a virus.

XX Sequence 3185 BP; 676 A; 883 C; 945 G; 681 T; 0 other;

XX Query Match

XX Best Local Similarity 31.1%; Score 790.4; DB 24; Length 3185;

XX Pred. No. 7.8e-181;

	Matches	855;	Conservative	0;	Mismatches	56;	Indels	5;	Gaps	2;
QY	4	CGCGTCCGGCCGCGAGCCGCGGCTAGAGCCCGCGGCGGCTCTAGCCCGAGCGCGCGCGTGT	63							
DB	52	CGCATGAAGCGCGAGCGCGCGGCTAGAGCCCGCGGCGGCTCTAGCCCGAGCGCGCGCGCG	111							
QY	64	GAGGGCCGATCCCGGC-CCCGGCTCCGGTTCCCGGGCGGCGGCGGCTGCTACCATGC	122							
DB	112	GGCGCTGGGCTTGGGCTCCCGGCTCCCGGCTCCCGGCTCCCGGCTCCCGGCTCCCGGCT	171							
QY	123	CGGCAAGCACCAGCACTTCCAGGAACCCGAGTCCGCTCTCGCGGAAATACCTTCCTGT	182							
DB	172	CCGCAAGCACCAGCACTTCCAGGAACCTGAGTCCGCTCCCGGAAATACCTTCCTGT	231							
QY	183	TTGGCTTCAACATGTTTCTGGTCTCGGAGCCCTGTTCTCGGCAATCGGCTCTGGG	242							
DB	232	TTGGCTTCAACATGTTTCTGGTCTCGGAGCCCTGTTCTCGGCTATCGGCTCTGGG	291							
QY	243	CTGGGGTGAGAGGGTCTCTCCAACTCTCGGCTGACCGATCTGGAGGCGCTCG	302							
DB	292	CTGGGGTGAGAGGGGCTTCTCTCAACATCTCAGCGCTGACAGATCTGGAGGCGCTG	351							
QY	303	ACCTCTGTGGCTGTTTGTAGTGGTGGAGGCGTCAATGCTCGTGGGCTTTGCCGCT	362							
DB	352	ACCCGCTGTGGCTGTTTGTAGTGGTGGAGGCGTCAATGCTCGTGGGCTTTGCTGCT	411							
QY	363	GCATCGGGCTCTCCGGGAGAACATTTCTGCTCAAGTTTCTCAGTGTTCCTTGCC	422							
DB	412	GCATTTGGGCGCTCCCGGAGAACACCTTCTCTCAAGTTTCTCGTGTTCCTGCTG	471							
QY	423	TCATCTTCTCTGGAGCTGGCAACAGGATCTTGGCTTCGTTATTCAGGACTGGATTC	482							
DB	472	TCATCTTCTCTGGAGCTGGCAACAGGATCTTGGCTTCGTTATTCAGGACTGGATTC	531							
QY	483	GAGACGAGCTCAATTTCTTCAATTAACAACAACGCTCAAGGCTATCGGATGACATTGACC	542							
DB	532	GAGACGAGCTCAACCTTCTCATCAACAACAACGCTCAAGGCTATCGGATGACATTGACC	591							
QY	543	TCAGAACTCATTTGATTTGCTCAGGAATATTTGGTCTTGTCTCGGAGCCCGAGGCGCTA	602							
DB	592	TCAGAACTCATTTGATTTGCTCAGGAATATTTGGTCTTGTCTCGGAGCCCGAGGCGCTA	651							
QY	603	ATGACTGGAACCTCAATATCTATTTCAATGCACTGACTTGAACCCGAGCCGAGAGCGCT	662							
DB	652	ATGACTGGAACCTCAATATCTATTTCAATGCACTGACTTGAACCCGAGCCGAGAGCGCT	711							
QY	663	GGGGGTGCGCTTCTCTGCTGTGTCAGGAGCCCTGC---GATGCTCTCAACACCCAGT	718							
DB	712	GGGGGTGCGCTTCTCTGCTGTGTCAGGAGCCCTGCAGGAGATGCTCTCAACACCCAGT	771							
QY	719	GTGGCTATGATGTCGGCTCAAACTGGAGCTGGAGCAGAGGCTCCATACACACCAAG	778							
DB	772	GTGGCTATGATGTCGGCTCAAACTGGAGCTGGAGCAGAGGCTCCATACACCAAG	831							
QY	779	GCTGTGTGGGCGAGTTGAGAGTGGCTGCAGAGCAACCTGATCGTGGTGGGCTCT	838							
DB	832	GCTGTGTGGGCGAGTTGAGAGTGGCTGCAGAGCAACCTGATCGTGGTGGGCTCT	891							
QY	839	TTGTGGGCACTGCTCTCTCCAGATCTTGGTATCTGCTGGCCCAAGAACCTTGTGAGTG	898							
DB	892	TCATGGGCACTGCGCTCTCTCCAGATCTTGGGATCTGCTGCTGCCCAAGAACCTTGTGAGTG	951							
QY	899	ACATCAAGGCACTGAAGCCAACTCG	924							
DB	952	ACATCAAGGCACTGAAGCCAACTCG	977							

## RESULT 5

AAC90014

ID AAC90014 standard; cDNA; 1126 BP.

XX

XX AAC90014;

XX



09-MAR-2001 (first entry)  
Clone HNTMH27 coding sequence.  
Gene therapy; human; 4 transmembrane superfamily receptor protein; endocrine; cardiovascular; cerebrovascular disease; neural disorder; reproductive; skin; renal system; autoimmune; hyperproliferative; ocular; bacterial infection; viral; fungal; ss.  
Homo sapiens.  
W0200070076-A1.  
23-NOV-2000.  
18-MAY-2000; 2000WO-US13504.  
19-MAY-1999; 99US-0135122.  
03-JUN-1999; 99US-0137797.  
11-JUN-1999; 99US-0138573.  
18-AUG-1999; 99US-0149447.  
28-JAN-2000; 2000US-0178770.  
(HUMA-) HUMAN GENOME SCI INC.  
Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA; Rosen CA;  
WPI: 2001-007502/01.  
P-PSDB; AAB49504.  
Isolated nucleic acid molecule encoding human soluble 4 transmembrane superfamily receptor protein, useful for diagnosing, treating and/or preventing disorders e.g. Alzheimer's, cancer and arrhythmia -  
Claim 1; Pages 269-270; 297pp; English.  
The present invention relates to isolated nucleic acids and proteins encoding human soluble 4 transmembrane superfamily receptor protein (see AAC90012-C90023 and AAB49502-B49513). The present sequence is one such nucleic acid. The present sequence is useful for preventing, treating or ameliorating a medical condition and in diagnosing (susceptibility to) a pathological condition e.g. endocrine disorders e.g. Addison's disease, (cardio)vascular diseases e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural disorders e.g. Alzheimer's and Parkinson's disease, reproductive disorders e.g. skin disorders e.g. psoriasis, renal system disorders e.g. nephritis (auto)immune system disorders e.g. graft vs. host disease, hyperproliferative disorders e.g. neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections caused by bacteria, viruses and fungi.  
Sequence 1126 BP; 191 A; 352 C; 343 G; 240 T; 0 other;  
Query Match 31.0%; Score 787.2; DB 22; Length 1126;  
Best Local Similarity 92.8%; Pred. NO. 3.1e-180;  
Matches 850; Conservative 0; Mismatches 63; Indels 5; Gaps 2;  
12 GCGCAGCGCCGGGCTAGGCCCCCGGGCGCTCTAGCCAGGCGCGCGGTGGAGGGCG 71  
49 GCGCAGCGCCGGGCTAGGCCCCCGGGCGCTCTAGCCAGGCGCGCGGTGGAGGGCG 108  
72 ATCCCGGC -CCCGGCTCCGGTTCCCGGCGCGCGCGGTGTCTACCATGCGGGGCAAG 130  
109 GCGCTGGCTCCGGCTCCGGTTCCCGGCGCGCGGTGGCGGTCTACCATGCGGGGCAAG 168  
131 CACGAGCACTCCAGAACCGAGGTGGCTGGTGGGAAATCTTCTGTTGGCTTC 190  
169 CACGAGCACTCCAGAACCTGAGGTGGCTGGTGGGAAATCTTCTGTTGGCTTC 228  
191 AACATTGTTTCTGGGTGCTGGAGCCCTGTTCTGGCCATCGGCTCTGGGCGTGGGT 250  
229 AACATTGTTTCTGGGTGCTGGAGCCCTGTTCTGGCTATCGGCTCTGGGCGTGGGT 288

QY 251 GAGAAGGGTGTCTCTCCAAACATCTCTGCGCTGACCGATCTGGGAGGCGCTCGACCCCTGTG 310  
Db 289 GAGAAGGGCGTTCTCTCGAAACATCTCAGCCCTGACAGATCTGGGAGGCGCTTGAACCCGTG 348  
QY 311 TGGCTGTTTGTAGTGGTTGGAGGCGTCAATGTCGGTGTCTGGGCTTTGGCGGTGATCGGG 370  
Db 349 TGGCTGTTTGTAGTGGTTGGAGGCGTCAATGTCGGTGTCTGGGCTTTGGCTGATTTGGG 408  
QY 371 GCTCTCGGGGAGAACACTTTCTCTGCTCAAGTCTTCTCAGTGTCTCTTGGCTCATCTTC 430  
Db 409 GCGCTCGGGGAGAACACTTTCTCTGCTCAAGTCTTCTCTCGGTCTCATCTTC 468  
QY 431 TTCTCTGGAGCTGGCAACAGGCGATCTTGGGCTTCTGATTTCAAGGACTCGATTCGAGACCG 490  
Db 469 TTCTCTGGAGCTGGCAACAGGCGATCTTGGGCTTCTGATTTCAAGGACTCGATTCGAGACCG 528  
QY 491 CTCATTTCTTCTTAAACAAACAGTCAAGGCGCTATCGGGATGACATTTGACCTCCAGAAC 550  
Db 529 CTCATTTCTTCTTAAACAAACAGTCAAGGCGCTATCGGGATGACATTTGACCTCCAGAAC 588  
QY 551 CTCATTTGACTTTGCTCAGGAATATTGCTTCTGCTGGAGCGCGGAGGCGCTTAATGACTGG 610  
Db 589 CTCATTTGACTTTGCTCAGGAATATTGCTTCTGCTGGAGCGCGGAGGCGCTTAATGACTGG 648  
QY 611 AACCTCAATATCTATTTTCAACTGCACTTGAAACCGAGCGGAGCGCTGCGGGGTG 670  
Db 649 AACCTCAATATCTATTTTCAACTGCACTTGAAACCGAGCGGAGCGCTGCGGGGTG 708  
QY 671 CCCTTCTCTGCTGTGTCTAGGGACCTTGC - - - - - GATGTCTCTCAACACCCAGTGTGGCTAT 726  
Db 709 CCCTTCTCTGCTGTGTCTAGGGACCTTGC - - - - - GATGTCTCTCAACACCCAGTGTGGCTAT 768  
QY 727 GATGTCGGCTCAACTGAGCTGGAGCGAGCGGCTCAATACACCAAGGCTGTGTG 786  
Db 769 GAGTCCGGCTCAACTGAGCTGGAGCGAGCGGCTTATCCACACCAAGGCTGTGTG 828  
QY 787 GGCAGTGTGAGAAGTGGTGCAGGACAACTGATGCTGTGGTGGCTGGGCTCTTGTGGGC 846  
Db 829 GGCAGTGTGAGAAGTGGTGCAGGACAACTGATGCTGTGGTGGGAGTCTTATCATGGC 889  
QY 847 ATGCTCTCTCTCAGATCTTTGGTATCTGCTGGCGCCAGAACCTTGTGAGTGACATCAAG 906  
Db 889 ATGCGCTCTCTCAGATCTTTGGCATCTGCTGGCGCCAGAACCTTGTGAGTGACATCAAG 948  
QY 907 GCAGTGAAGGCCCACTGG 924  
Db 949 GCAGTGAAGGCCCACTGG 966  
RESULT 6  
ABZ11816  
ID ABZ11816 standard; cDNA; 1023 BP.  
XX  
AC ABZ11816;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 698.  
XX  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cystostatic; immunomodulator; neurotropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.  
OS Homo sapiens.  
XX  
PN W0200270539-A2.  
XX  
PD 12-SEP-2002.

XX	05-MAR-2002; 2002WO-US05095.	
PF	05-MAR-2001; 2001US-0799451.	
XX	(HYSE-) HYSEQ INC.	
XX	Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;	
PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;	
PI	Wehrman T, Wang J, Wang D, Drmanac RT;	
XX	WPI; 2002-759812/82.	
DR	P-FSDB; ABP69599.	
DR		
XX	New polynucleotides comprising sequences assembled from expressed	
PT	sequence tags (ESTs), useful for treating cell-proliferative,	
PT	neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or	
PT	platelet or coagulation disorders	
XX	Claim 1; SEQ ID NO 698; 1012pp + Sequence Listing; English.	
PS	The invention relates to an isolated polynucleotide (I) comprising a	
XX	nucleotide sequence selected from any of 948 sequences	
CC	(ABZ11119-ABZ12066) or their mature protein coding portion, active domain	
CC	coding protein or complementary sequences. The polynucleotides are useful	
CC	for identifying expressed genes or for physical mapping of human genome.	
CC	The encoded polypeptides (ABP68902-ABP69849) are useful as molecular	
CC	weight markers, as a food supplement, for generating antibodies, in	
CC	medical imaging, screening and diagnostic assays and for treating	
CC	cell-proliferative disorders (cancer), neurodegenerative diseases	
CC	(Parkinson's or Alzheimer's disease), autoimmune diseases (multiple	
CC	sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid	
CC	disorders, platelet or coagulation disorders, wound, burns, incision,	
CC	ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,	
CC	parasitic), arthritis, etc.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 1023 BP; 190 A; 300 C; 293 G; 240 T; 0 other;	
	Query Match 30.5%; Score 773.4; DB 24; Length 1023;	
	Best Local Similarity 88.1%; Mismatches 0; Mismatches 177;	
	Matches 922; Conservative 0; Pseudomatches 91; Indels 34; Gaps 6;	
Qy	89 GGTTCCTCCGGGCGGGCGGCGTGTCTACACATGCCGGCAAGCACGACATTCACAGAA 148	
Db	1 GGTTCCTCCGGGCGGGCGGCGTGTCTACACATGCCGGCAAGCACGACATTCACAGAA 60	
Qy	149 CCCAGAGTCGGCTGTCTCGGGGAAATATCTCTGTTTGGCTTCAACATTTTCTCGGGTG 208	
Db	61 CCTGAGGTGGCTGTCTCGGGGAAATATCTCTGTTTGGCTTCAACATTTCTCTCGGGTG 120	
Qy	209 CTGGAGCCCTGTTCTTGGCCATCGGCTCTGGCCCTGGGGTGGAGGGTGTCTCTCC 268	
Db	121 CTGGAGCCCTGTTCTTGGCTATCGGCTCTGGCCCTGGGGTGGAGGGGCTTCTCTCG 180	
Qy	269 AACATCTCTGGCTGACCGATCTGGGAGCCCTCGACCCCTGTGTGCG-TGTTTGTAGTCGT 327	
Db	181 AACATCTCAGGCTGACAGATCTGGGAGCCCTTGACCCCTGTGTGTGTGTGTGTGTGTGT 240	
Qy	328 TGGAGGGCTCATGTCCGTGTGGGGCTTTGCGG--CTGCAATCGGGGCTCTCCGGGAGAAC 385	
Db	241 TGGAGGGCTCATGTCCGTGTGGGGCTTTGCTGGGCTGCAATGGGGCCCTCCGGGAGAAC 300	
Qy	386 ACTTTCCTGCTCAAGTTTCTCAGTGTTCCTTGGCTCATCTTCTTCCTGGAGCTGGCA 445	
Db	301 ACCTTCTGCTCAAGTTTCTCAGTGTTCCTTGGCTCATCTTCTTCCTGGAGCTGGCA 360	
Qy	446 ACAGGGATCTTGGCTTTCGTATTCAAGGACTGGATTGAGACACGAGTCAATTTCTTCATT 505	
Db	361 ACAGGGATCTTGGCTTTCGTATTCAAGGACTGGATTGAGACACGAGTCAATTTCTTCATT 420	

XX PR 18-JUN-1999; 99FR-0007784.  
 XX PA (UYPA-) UNIV PASTEUR LOUIS.  
 XX PI Andriamanpandry C, Maitre M;  
 XX DR P-PSDB; ABP98695.  
 XX PT New rat brain gamma-hydroxybutyrate receptor and its encoding nucleic  
 XX PT acid, useful for identifying agents for treating e.g. epilepsy -  
 XX PS Claim 5; Page 64-66; 66pp; French.  
 XX CC This sequence represents the cDNA sequence encoding a novel rat  
 CC gamma-hydroxybutyrate receptor (GHR) isolated from rat brain hypocampal  
 CC cell cDNA library. The invention relates to the isolation of this novel  
 CC sequence, fragments of it and homologues of the sequence except for those  
 CC homologues that having GenBank accession numbers AAC 17120 (human  
 CC tetraspan NET-4), AA615405, AA967250 or A1467230. GHR is the receptor  
 CC for gamma-hydroxybutyrate in the rat brain and is involved in regulating  
 CC dopaminergic, opioid and GABA (gamma-aminobutyric acid)ergic activities.  
 CC The nucleic acid that encodes the protein is used: (1) as primers or  
 CC probes for detection/amplification, particularly for screening gene  
 CC libraries to identify promoters and regulators of the GHR gene; (2) for  
 CC expression of recombinant polypeptides; and (3) to detect allelic  
 CC variants, mutations, deletions, loss of heterozygosity or genetic  
 CC abnormalities in the GHR gene (for diagnosing diseases, or  
 CC susceptibility, associated with abnormal expression of GHR). The  
 CC protein (or cells and transgenic animals expressing it) is used:  
 CC (1) to screen for agents that interact with GHR; (2) to study  
 CC expression/activity of the receptor, including its interaction with  
 CC other compounds; and (3) to raise antibodies (Ab) specific for GHR.  
 CC The Ab are used to detect/measure (I) in diagnostic immunoassays.  
 CC The agents are used to prevent or treat diseases associated with abnormal  
 CC expression or activity of GHR, particularly those involving cerebral  
 CC GABA (gamma-aminobutyric acid)ergic and/or dopaminergic activities, e.g.  
 CC epilepsy, anxiety, sleep or behavioral disorders, withdrawal from  
 CC addictive drugs, neurodegeneration (Parkinson's disease), psychoses,  
 CC schizophrenia and regulation of secretion of hormones (growth hormone  
 CC and prolactin) that are under dopaminergic control.  
 XX SQ Sequence 1567 BP; 281 A; 436 C; 480 G; 370 T; 0 other;  
 Query Match 20.1%; Score 511; DB 23; Length 1567;  
 Best Local Similarity 85.9%; Pred. No. 2.3e-113;  
 Matches 602; Conservative 0; Mismatches 95; Indels 4; Gaps 3;  
 QY 12 GCCGAGCGCGCGGCTAGGCCCGCGGGCTCTAGCCCGCGCGCGCGTGGAGGCGG 71  
 DB 28 GCCGAGCGCGCGGCTAGGCCCGCGGGCTCCAGCGCTGGCGAGCGCGAGGGCTG 87  
 QY 72 ATCCCGCGCGCGGCTCCGCGCGCGCGCGGGCGGCTGCTACCATGCCGCGCAAGC 131  
 DB 88 GCGCGGTTCCCGGCTCGGTTCCAGCGCGCGAGGTGGCGCTCACCATGCCCGCAAGC 147  
 QY 132 ACCAGCACTCCAGCAACCCGAGGTGCGCTGCTGCGGGAAATCTCTGTTGGCTTCA 191  
 DB 148 ATCAGCAATCCAGACCCCTGAGGTGGCTGCTGCGGGAAATCTCTGTTGGCTTCA 207  
 QY 192 ACATGTTTTCTGGTGTGGAGCCCTGTTCTCGGCCATCGGCTCTGGGCTGGGGTG 251  
 DB 208 ACATGTTCTTGGTGTGGAGCCCTGTTCTCGGCCATCGGCTCTGGGCTGGGGTG 267  
 QY 252 AGAGGGTGTCTCTCCACATCTCTGGCTGACCGAT-CTGGGAGGCTCGACCTGTG 310  
 DB 268 AGAAGGGCGTTCTTTTCCAACTCTCAGGGCGGACAGATCCAGGCGGTCTTGACCCCTG 327  
 QY 311 TGGCTGTTGTAGTGGTGGAGGCTCATGTCGCTGGCTGGCTGGCTGGCTGGCTGG 370  
 DB 328 TGGCTGTTGTAGTGGTGGAGGATCATGTCAGTGGCTGGCTGGCTGGCTGGCTGG 387

QY 371 GCTCTCCGGAGAACACTTTCTGCTCAAGTTTTTCTCAGTGTCTTGGCCTCATCTTC 430  
 DB 388 GCCTCCGGGAAACACCTTCTGCTCAAAATTTTCTGTGTCTTCGGCCTCATCTTC 447  
 QY 431 TTCTGGAGCTGGCAACAGGGATCTTGGCCTTGTATTTCAAGGACTGGATTCGAGACCAG 490  
 DB 448 TTCTGGAGCTGGCGCGGGATCTTGGCCTTGTGTTCAGGATTTGGATCCGAGACAA 507  
 QY 491 CTCATTTCTTCTTATTAACAACAGCTCAAGGCTATCGGGATGACATTTGACCTCCAGAAC 550  
 DB 508 CTTAACTCTTCTCATCAACAACAAATGTCAAAGCCTACCGGAGCATATTTGACCTTCAGAAC 567  
 QY 551 CTCATTTGACTTTGCTCAGGAATATTTGCTTGTCTGGAGCCCGAGGCTCTAATGACTGG 610  
 DB 568 CTTATGCACTTTGCTCAGGAATATTTGCTTGTCTGGAGCCCGAGGCTCAATGACTGG 627  
 QY 611 AACCTCAATATC-TATTTCAACTGCATGACTT-GAACCAGCGAGAGCGCTCGCGG 667  
 DB 628 AACCTCAACATCGGACTTCAACTGCATGACTTCAAAACCCAGCGCGAGCGCTGTGGG 687  
 QY 668 GTGCCCTTCTCTGCTGTGTCTGAGGACCTCTGCGATGTCTC 708  
 DB 688 GTGCCCTTCTCTGCTGTGGGTAAAGACCTCTGCGGAAGACGTC 728  
 RESULT 8  
 AAH98169  
 ID AAH98169 standard; cDNA; 1735 BP.  
 XX AC AAH98169;  
 XX DT 12-OCT-2001 (first entry)  
 XX DE Human EST-derived coding sequence SEQ ID NO: 26.  
 XX KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; Gene therapy; nutrition; SS.  
 XX US Homo sapiens.  
 XX PN WO200154477-A2.  
 XX PD 02-AUG-2001.  
 XX PF 25-JAN-2001; 2001WO-US02687.  
 XX PR 25-JAN-2000; 2000US-0491404.  
 PR 17-JUL-2000; 2000US-0617746.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 15-SEP-2000; 2000US-0663870.  
 XX (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
 XX WPI; 2001-476164/51.  
 DR P-PSDB; AAM23510.  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 PS Claim 1; Page 200-201; 1275pp; English.  
 XX The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA

CC of the invention.

XX Sequence 1735 BP; 363 A; 463 C; 528 G; 381 T; 0 other;

SQ Query Match 17.7%; Score 449.2; DB 22; Length 1735;

Best Local Similarity 64.6%; Pred. No. 2.1e-98;

Matches 1106; Conservative 0; Mismatches 473; Indels 133; Gaps 24;

QY 732 CCGGCTCAAACTGGAGCTGAGCAGCAGAGGCTCCATACACACCAAGGCTGTGGGCCA 791  
DB 1 CCGGCTCAAACTGGAGCTGAGCAGCAGAGGCTCCATCCACCAAAAGGCTGTGGGCCA 60  
QY 792 GTTTGAGAAGTGGCTGCAGGACAACTGATCGTGGTGGCTCTTTGTGGGCATCGC 851  
DB 61 GTTTGAGAAGTGGCTGCAGGACAACTGATTTGTGGTGGGAGTCTTTCATGGGCATCGC 120  
QY 852 TCTCTCCAGATCTTTGGTATCTGCTGGGCCAGAACCTTTGTGAGTACATCAAGGAGT 911  
DB 121 CTTCTCCAGATCTTTGGCATCTGCTGGGCCAGAACCTCGTGAATGACATCAAGGAGT 180  
QY 912 GAAGGCCAACTGGATCAAACTGATGATGCTCAAACTACTCAAAATAAACAACCTTG 971  
DB 181 GAAGGCCAACTGGAGCAAAATGGAATGATG-----ACTTTG 215  
QY 972 AAAACCACTGGCTTACGCCCAACCATCTCAGAGGTTCCATGGGCCGAGGCCCTCAGCCGT 1031  
DB 216 AAAACCACTGGCTTACGCCCAACCATTTCCGAGGTCCTGTCCACGGCGGGCCCTCAGAGA 275  
QY 1032 GCGCTCTGCTGGGGCCCC-AGGCCAGACCCACCTCGGCCAACATGTTTTC-TTGGCCTGG 1089  
DB 276 ACTCTCTGACTGGGGCCCCGGGGCCCCAGCCACCCAGCCAGACATGTTTCTTTGGGCTGG 335  
QY 1090 GTAGTACATACGATGAGCCAACTTTAAACTTTGGCATTTTCAATGTAAGTCCAGATC 1149  
DB 336 GTGGTTTATACCTGAGCCAACTTTTAAAAATTTGGTAGATTCACATAAAGTCCAGATC 395  
QY 1150 CCAGCATCTTGTGAAGAATGGCCATCCGGCCACAGCGGCTCTTCTATGGCTTCGTCTCC 1209  
DB 396 CACAGCTCTCTTGAAGAATGACACCTGGCTACGGCGGCTCTTCGTGTCACACTAACC 455  
QY 1210 TGGGATGGCGCTTCTGTTCTCTGAGGACCCACCTCAACCGTGTCC-----TG 1260  
DB 456 TGGGACACTGCTCCCTCCAGTCAACAGGGGCCAGCTGGCCGCTTCTACTCACTAAGTG 515  
QY 1261 CTTGCTGACCTGGAGGCTGGAGCTGGCTCTCTCCACCTCTGCAAGTTTTCCTCTG 1320  
DB 516 CCGCTGACCTTGTACACTAGAGCTGGCTC-CCACCTCTGAGGGTTATTTCTCTG 573  
QY 1321 AAATGCTGAAGCTGCTGTGGGCCAAGCCCGGATCGAAGCTGGAGCGTGAAGAATTGG 1380  
DB 574 ACCT----CGAGGCGCTGGGGCCAACTCTGGAGTGAACACAGGGGACCTGAAGGA-TGG 628  
QY 1381 GGAGGCTGGAGCTGCCCCAAGAGGCCACAGCTGGGAAGGCTGTGGCCCTCTGGGGG 1440  
DB 629 AGAGGCTGGAGCCCGCTTTGAAGAGGCTGAGCTGGAGGG-CCGGCTTGTCTGGGA 686  
QY 1441 CCAAGATGGGTGCCACCGTCCAGAGAGTGGCCGGAGGGTGGGATGAGATCAAGGAAG 1500  
DB 687 CTCGGTGGAG-TAGATGCTCCAGAGAGGCTCTGAGGGGTGGATGGGGTTCAGACA 745  
QY 1501 GTTTGGGCAAGACGTAGCTGGAAGCTGAGCTTGTGATCCCATGGGGATGGGAGAGCC 1560  
DB 746 ATTTTGAAGAAGTAGCTGGAAGCC-----ATGGGACTGGCGGGAGCC 790  
QY 1561 TGTTGAGGCGGCTGATGTGTAGACTCAGCTCTGTTGGAATCAGTTCAAAATCTTC 1620  
DB 791 TGTTTGGGATCTGATGTTGATGTTGACTCC-----TAGGAGTCAAGTTTCAGCATCTTCG 842  
QY 1621 CAGTGGCTGTAGATTGCTCTGACCACTAGAGGGCGGCCACACAGCATTCCTCG 1680  
DB 843 CCGTGG-CTGCAGAGCTGCTGATGGCACTAGAGGGCAGCCAGCCACACTCCTCG 901  
QY 1681 GTCTGCTTCTTAGGACAAACCCCACTAGTACAGCCCTGTGCTGGTGTGTCTCACCTG 1740

DB 902 GTCTGGCTTCTCCCGCAACCTCACTCTAGTAGAGCTG-----TGCTG 946  
QY 1741 CTTACTAGTTCCTTTGGGTTTCATGGAATTTACAAGCTTCTTAAAGGACAGAGTGGCTCAG 1800  
DB 947 CTTACTAGGCTCTGGGTTTCGAGAGATTTGGGAATTTCTCA--GAGCCAACTGGCTCAG 1004  
QY 1801 ATTGGGAAGCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGTGTGTGGAGTATT-- 1858  
DB 1005 GCTTGGGAAGGCTGGCTGCTGCCCTCAGCTCCGCTCATCAGCTATGTGAAGGGGTGTGT 1064  
QY 1859 -----TGTAATCAAAAGGAGAGGTTTGGCTAGTGGCCAGTCTTTTAACTTAGATGCCCC 1912  
DB 1065 ATGAGTGTATCTGCCGCCCTCCCTGGCTGGTCCAGAGATCTCAAACTCCGATGCCCC 1124  
QY 1913 TCAGGCCCGGGTGGGTTATAAAA-----ATAAAGTAGGCTTTGAGCTGTGAGGCCCTTTGG 1968  
DB 1125 CTGGGCCCACTGATGTTGTAAATGGAATGAACAGAGCCCTTGAAGTGGAGCTGCTTC 1184  
QY 1969 GACTTTAATTTTCCCACTATTCTGGAGATGGGACATAGAGAGACATTTGCTTGTGCTG 2028  
DB 1185 ACTTTGA--CTTTCCCACTGTTGCTGGAGACA-----AAGACATCGTGATG 1228  
QY 2029 AGAAATPACTTGCATGATTGAGTCTGAGTCTGAGTCCGCTAAGGGCAACTGGCTTGAATGACATCAA 2088  
DB 1229 AGAAGAGTTTCGACAAT-----CTAGTCGGTAAACAGCCACTTTCCTTGAGACCAAGAGA 1283  
QY 2089 GGGGTGGTGGGACTGTGGCAACACAGATTTCCCACTGAAATTTGGTGGCTGTC--CTT 2146  
DB 1284 GTGCGGTGGGATGGGGGGAGAGCACGGGTCCCGTCTGACAGTGGCGGCTGCCATATT 1343  
QY 2147 CGTGTGGGCTAATGGCTGACAGCGAGAACTGTAGTGGTCTGCTCTAATGGGAGAGTC 2206  
DB 1344 CAGGTGTAGCTAATTTGCTCTGGTGGGAATGCAG-----GCCTAATGACAGAAATC 1395  
QY 2207 TGGAGAAGCAAGAAGCTAGATTTTTCATGTGAATATCCGAGTTTAAAGTTGTTGCA 2266  
DB 1396 TGGAGAAGCCAGAAATACAGATTTGT-ATGTGAGATGCTCTGATTTTAAAGTTTGGC 1454  
QY 2267 GCTAATGAGAAAACCTCTTAAACCTGATATGCAAAAGGTGTGGGGCCATCTTTGACA 2326  
DB 1455 AGAAATTAATTCAGAAATCAAAATTCAGAGCCCAACAGAGTGCAGGACCCAGCTTTGGCC 1514  
QY 2327 CTTCCCCCACCATAGTCCCTCAGGACAGAGTCCCATGGAGCCCTGGTGGTGCACCG 2386  
DB 1515 CCATGCCC-CTGTAGTCCCTCTGGGACAGTCAACGCTGGG-GTCTGGCTGCTGTGCA 1572  
QY 2387 TTGAGGAATGCTGGGAATGCTGTCAGGTGGG 2418  
DB 1573 TTGAGGATGCTGGGACATGCTGCCGGGTGGC 1604

## RESULT 9

AAC98195

ID AAC98195 standard; cdna; 1655 bp.

XX AAC98195;

AC AAC98195;

XX 09-MAR-2001 (first entry)

XX Human colon cancer antigen nucleotide sequence SEQ ID NO:205.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX identification; cytostatic; cardiotoxic; neuroprotective; vulnary;

XX immunomodulatory; muscular; gynaecological; gastrointestinal;

XX nephrotropic; anti-infective; antibacterial; gene therapy; wound;

XX neural disorder; immune system disorder; muscular disorder;

XX reproductive disorder; gastrointestinal disorder; renal disorder;

XX infectious disease; cardiovascular disorder; 88.

XX Homo sapiens.

XX WO20005351-A1.



CC nucleic acid. The present sequence is useful for preventing, treating or ameliorating a medical condition and in diagnosing (susceptibility to) a pathological condition e.g. endocrine disorders e.g. Addison's disease, (cardio)vascular diseases e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural disorders e.g. Alzheimer's and Parkinson's disease, reproductive disorders, skin disorders e.g. psoriasis, renal system disorders e.g. nephritis, (auto)immune system disorders e.g. graft vs. host disease, hyperproliferative disorders e.g. neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections caused by bacteria, viruses and fungi.

XX Sequence 1174 BP; 306 A; 259 C; 291 G; 316 T; 2 other;

Query Match 16.5%; Score 418.6; DB 22; Length 1174;  
Best Local Similarity 73.0%; Pred. No. 4.6e-91;  
Matches 580; Conservative 0; Mismatches 209; Indels 6; Gaps 3;

QY 135 AGCACTTCCAGGAACCCGAGCTGGCTGCTGCGGGAATACCTTCTGTTGGCTTCAACA 194  
DB 103 AGCACTACAGGGTCTGAGTCACTGTTGTCATCAATCTTCAATTTGGCTTCAATG 162  
QY 195 TTGTTTCTGGTCTGGAGCCCTGTTCTGCGCATCGGCCCTGCGGCTGGGGTGAGA 254  
DB 163 TCATATTTGGTTTTTGGGAATAACAATTTCTTGAATTTGACTGTGGCATGGAATGAAA 222  
QY 255 AGGTGTTCTCCACATCTCTGCGTGAACGATCTGGAGGCTCGACCCCTGTGTGGC 314  
DB 223 AAGGAGTTCTGCCAATCTCTTCCATCACCGATCTCGCGGCTTTTGACCCAGTTTGGC 282  
QY 315 TGTGTTGAGTGGTGGAGGCTCATGTCCGTGCTGGCTTTGCGGCTGCATCGGGCTC 374  
DB 283 TCITCTTGTGTTGGAGAGTATGTTCAATTTGGATTTGAGGTGATTTGGAGC-G 341  
QY 375 TCGGGAGAACACTTCTCTGCTCAAGTTTCTTCAAGTGTTCCTTGGCTCATCTTC-TTC 433  
DB 342 TACGGGAAACACTTCTCTCAAGTTTCTTCTGTTTCTGGGAAATTAATTTCTTTC 401  
QY 434 CTGAGCTGGCAACAGGATCTTGGCTTCGTATTCAGGACTGGATTCAGACCCAGCTC 493  
DB 402 CTGAGCTCACTCGCGAGTTCTAGCAATTTGTTTCAAGACTGGATTCAAAGACCAGCTG 461  
QY 494 AATTTCTTCAATTAACAAACGTCAGGCTATCGGGATGACATTTGACCTCCAGAACCTC 553  
DB 462 TATTTCTTATAAACAACATCAGAGCATATCGGGATGACATTTGATTTGCAAAACCTC 521  
QY 554 ATTGACTTCTCAGGAATATTGTTCTGTCGGAGCCCGAGGCTTAATGACTGGAAC 613  
DB 522 ATAGACTTCAACCCAGGAATATTGGCAGTGTCTGTTGGGCTTTTGGAGCTGATTTGGAAC 581  
QY 614 CTCAATATCTATTCAACTGCACTGACTTGAACCCGAGCCGAGCGCTGCGGGTGGCC 673  
DB 582 CTAAATATTTACTTCAATTCACAGATTCCAAATGCAAGTCGAGAGCAATGTCGCTTCCA 641  
QY 674 TTCTCTGCTGTGTACGGGACCTGTC---GATGTCTCTCAACCCAGTGTGCTGATGAT 729  
DB 642 TTCTCTGCTGTCTAAAGATCCCGCAGAGATGTCTCAACTCAACTCAGTGTGCTATGAT 701  
QY 730 GTCCGCTCAACTGGAGCTGGAGCAGCGGCTCATACACACCAAGCTGTGTGGC 789  
DB 702 GCCAGGCAAAACCCAGAGTTGACCAAGTATGTAATCTACACGAAAGCTGTGTGCC 761  
QY 790 CAGTTTGAGAGTGGCTGAGGACCACTCATCTGTGTGCTGGCTGCTTTGTGGGCATC 849  
DB 762 CAGTTTGAGAGTGGTTCAGGACAATTTAACCATCTGTGCTGTAATTTTCATAGGCAT 821  
QY 850 GTCTCTCTCAGATCTTTGGTATCTGCTGGCCAGAACCTTGTGAGTGACATCAAGGCA 909  
DB 822 GCATTGCTGCAGATATTTGGGATATGCTGGCCAGAAATTTGTTAGCGATATCGAGCT 881  
QY 910 GTGAGGCCAATCGG 924  
DB 882 GTCAGGGCGAGCTGG 896

RESULT 11  
AAC90020

ID AAC90020 standard; cDNA; 1178 BP.

XX AAC90020;

XX AC AAC90020;

DT 09-MAR-2001 (first entry)

XX DE Clone HB8EL16 coding sequence #2.

XX KW Gene therapy; human; 4 transmembrane superfamily receptor protein;  
XX KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
XX KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
XX KW bacterial infection; viral; fungal; ss.

OS Homo sapiens.

XX WQ200070076-A1.

XX 23-NOV-2000.

XX 18-MAY-2000; 2000WO-US13504.

XX 19-MAY-1999; 99US-0135122.

XX 03-JUN-1999; 99US-0137797.

XX 11-JUN-1999; 99US-0138573.

XX 18-AUG-1999; 99US-0149447.

XX 28-JAN-2000; 2000US-0178770.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;

PI Rosen CA;

XX WPI: 2001-007502/01.

DR P-PSDB; AAB49510.

XX Isolated nucleic acid molecule encoding human soluble 4 transmembrane superfamily receptor protein, useful for diagnosing, treating and/or preventing disorders e.g. Alzheimer's, cancer and arrhythmia.

XX Claim 1; Pages 274-275; 297pp; English.

CC The present invention relates to isolated nucleic acids and proteins encoding human soluble 4 transmembrane superfamily receptor protein (see AAC90012-C90023 and AAB49502-B49513). The present sequence is one such nucleic acid. The present sequence is useful for preventing, treating or ameliorating a medical condition and in diagnosing (susceptibility to) a pathological condition e.g. endocrine disorders e.g. Addison's disease, (cardio)vascular diseases e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural disorders e.g. Alzheimer's and Parkinson's disease, reproductive disorders, skin disorders e.g. psoriasis, renal system disorders e.g. nephritis, (auto)immune system disorders e.g. graft vs. host disease, hyperproliferative disorders e.g. neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections caused by bacteria, viruses and fungi.

XX Sequence 1178 BP; 310 A; 261 C; 291 G; 316 T; 0 other;

Query Match 15.8%; Score 402.2; DB 22; Length 1178;  
Best Local Similarity 71.6%; Pred. No. 4.2e-87;  
Matches 571; Conservative 0; Mismatches 218; Indels 8; Gaps 3;

QY 135 AGCACTTCCAGGAACCCGAGTCTGCTGCGGGAATACCTTCTGTTGGCTTCAACA 194

DB 101 AGCACTACAGGGTCTGAGTCACTGTTGTCATCAATCTTCAATTTGGCTTCAATG 160

QY 195 TTGTTTCTGGTGTGGAGCCCTGTTCTCGGCCATCGGCTCTGGGGGTGAGA 254

DB 161 TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAAATGCGACTGTGGCATGGAATGAAA 220

QY 255 AGGTGTTCTCTCCAAACATCTCTGCGCTCGACCGATCTGGAGGCCCTCGACCTGTGTGCG 314



Db 221 AAGGAGTTCTGTCCAAATCTCTTCCATCCGATCTCGCGGCTTTGACCCAGTTTGGC 280  
Qy 315 TGTTGTAGTGGTGGAGGCTCATCTCCGTCTGGGCTTTGCGGCTGATCGGGGCTC 374  
Db 281 TCTTCTCTGTGGTGGAGGAGTGATGTTCATTTTGGGATTTGCAAGGTGATTTGGAGC-G 339  
Qy 375 TCCGGGAGAACACTTTCTCTGCTCAAGTTTCTTC- --AGTGTCTCTTGGGCTCATCTTCT 431  
Db 340 TACGGGAAACACTTCTCTCAAGTTTCTCTGTTCTCTCGGGAATATTTCTT 399  
Qy 432 TCTTGGAGCTGGCAACAGGATCTTGGCTTCTGTTTCAAGGACTGATTCGAGACGAGC 491  
Db 400 TCTTGGAGCTCACTGCGGAGTCTAGCATTTGTTTTCAAGACTGATCAAGACGAGC 459  
Qy 492 TCAATTTCTTCAATTAACAACTCAAGCTTATCGGATGATTCAGCTTCAGAAC 551  
Db 460 TGTATTTCTTTATATAACAACTACAGATATCGGATGATTCAGTTGTTGCAAAACC 519  
Qy 552 TCATTTGACTTTTGTCTCAGGAATATTTGGTCTTGGTGGAGCCGAGGCTTAATGACTGGA 611  
Db 520 TCATAGACTTCAACAGGAATATTTGGAGTCTGTTGGGCTTTTGGAGCTGATGATGGA 579  
Qy 612 ACCTCAATATCTATTTCAATGCACTGCACTTGAACCCGAGCCGAGAGCGTGGGGTGC 671  
Db 580 ACCTAAATATTTACTTCAATTTGCACAGATTCCTCAATGCAAGTCGAGAGCGATGTGGCTC 639  
Qy 672 CTTTCTCTCTGTGTGAGGAGCCCTGC- ---GATGTCTTCAACACCCAGCTGTGGCTATG 727  
Db 640 CATTTCTCTCTGCTCACTAAAGATCCCGCAGAGATGTCATCACTCACTGTGGCTATG 699  
Qy 728 ATGTCTCGGCTCAAACTGGAGCTGGAGCAGCAGGCTCTCATACACCAAAAGGCTGTGG 787  
Db 700 ATCCAGGCAAAACAGAGATTTGACAGCAGATTTGTAATCTACCAAAAGGCTGTGC 759  
Qy 788 GCGAGTTTGAAGTGTGCTGAGCAACCTGATCGTGGTGGGCTTTTGTGGCA 847  
Db 760 CCCAGTTTGAAGAGTGTGTCAGGACAAATTTAACCATCGTGTGCTGATTTTCATAGCA 819  
Qy 848 TCCTCTCTCTCCAGATCTTTGGTATCTGCTGCTGCCAGACCTTGTGAGTGACATCAGG 907  
Db 820 TTGCATTTGTCAGATATTTGGGATATGCTGCTGCCAGAAATTTGGTTAGCGATATCGAAG 879  
Qy 908 CAGTGAAGGCCAACTGG 924  
Db 880 CTGTGAGGCGAGCTGG 896

RESULT 12

ABK35735  
ID ABK35735 standard; cDNA; 1988 BP.

AC ABK35735;

DT 08-MAY-2002 (first entry)

DE cDNA sequence #126 encoding novel human secreted protein.

XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
XX immune deficiency disorder; blood disorder; inflammatory disorder;  
XX infectious disorder; allergic condition; neurodegenerative disorder;  
XX liver fibrosis; coagulation disorder; gene therapy; antimicrobial;  
XX tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.

OS Homo sapiens.

XX WO200177289-A2.

XX 18-OCT-2001.

XX 29-MAR-2001; 2001WO-US10232.

XX 06-APR-2000; 2000US-195605P.

XX (GENY ) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;  
PI Clark HF, Fechtel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;  
XX MPI; 2002-179322/23.

XX Six hundred and twenty three polynucleotides derived from a variety of  
human tissue sources which encode secreted proteins, useful for  
treating immune deficiencies and disorders such as autoimmune disorders

XX Claim 1; Page 146; 393pp; English.

XX The present invention relates to the isolation of novel cDNA sequences  
which encode human secreted proteins. The cDNA sequences have been  
derived from a variety of human tissues. The invention also provides  
a method for producing proteins from these polynucleotide sequences.  
The proteins are useful for identifying compounds that modulate their  
activity and production. The sequences of the invention are  
useful for treating diseases such as hyperproliferative disorders  
(e.g. cancer), immune deficiency disorders (e.g. severe combined  
immunodeficiency (SCID)), autoimmune disorders (e.g. multiple  
sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory  
disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),  
allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.  
Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.  
haemophilia), and tumours. The polynucleotide sequences of the  
invention are also useful in gene therapy. ABK35810-ABK36232 represent  
the cDNA sequences of the invention that encode for novel human  
secreted proteins.

XX Sequence 1988 BP; 503 A; 434 C; 446 G; 604 T; 1 other;

Query Match 14.6%; Score 370; DB 24; Length 1988;

Best Local Similarity 72.4%; Pred. No. 3.2e-79;

Matches 495; Conservative 0; Mismatches 185; Indels 4; Gaps 1;

Qy 245 TGGGGTGGAGAGGTTCTCTCCAAATCTCTGGCGTGCACCATCTGGGAGGCTCGAC 304

Db 2 TGGAGTGAAGAGGAGTCTGTCCAAATCTCTCCATCATCCCATCTCGGGGCTTTGAC 61

Qy 305 CTGTGTGGCTGTTTGTAGTGTGGAGGCTCATGTCCGTGTGGGCTTTTGGCGGTGC 364

Db 62 CCAGTTTGGCTCTTCTTGTGGTGGAGGAGTGATGTTCATTTTGGGATTTGAGGGTGC 121

Qy 365 ATCGGGGCTCTCGGAGAACACTTTCTGCTCAAGTTTTTCTCAGTGTTCCTTGGCCTC 424

Db 122 ATGGAGCGCTACGGGAAACACTTTCTTCTCAAGTTTTTCTGTGTCTCTGGGAAT 181

Qy 425 ATCTTCTCTCGGAGCTGCAACAGGATCTTGGGCTTGTGATTTCAAGGACTCGATTGCA 484

Db 182 ATTTTCTCTCGGAGCTCACTGCGGAGTCTAGCAATTTGTTTCAAGACTGGATCAA 241

Qy 485 GACCACTCAATTTCTTCAATTAACAAACGTCAGGCTTATCGGATGACATTGACCTC 544

Db 242 GACCACTGATTTTCTTATAAACAACATCAGAGCATATCGGATGACATTGATTG 301

Qy 545 CAGAACCTCATTTGACTTTCTCAGGAATATTTGTTCTGTGGAGCCCGAGGCTTAAT 604

Db 302 CAAAACCTCATAGACTTCAACCAGGAATATTTGCAAGTGTGTGGGCTTTTGGAGCTGAT 361

Qy 605 GACTGGAACCTCAATATCTATTTCACCTGACTTGAACCCGAGCCGAGAGCGCTGC 664

Db 362 GATTGGACCTAAATATTTACTTCAATTTGTCGCGAAGATGTCTAACACTAGTGT 421

Qy 665 GGGGTGCCCTTCTCTGTGTGTCTAGGACCCCTGC- ---GATGTCTCTAACACCCAGTGT 720

Db 422 GCGGTTCCATTTCTCTGTGTGCTAAAGATCCCGAGAGATGTCTATCACTCAGTGT 481

Qy 721 GGTATGATGTCCGGCTCAAACTGAGCTGGAGCAGAGGGCTTCATACACCAAGGC 780



Db 482 GGCTATATGCGAGGAAACCAAGAGTTGACGAGAGATTGTAATCTACACGAAGGC 541  
Qy 781 TGTGTGGGCCAGTTTGAAGAGTGGCTCGAGGACCAACCTGATCGTGGTGGCTGGGCTCTTT 840  
Db 542 TGTGTGCCCCAGTTTGAAGAGTGGTTCAGGACAAATTTAACCATCGTTGCTGGTATTTTC 601  
Qy 841 GTGGGCATCGCTCTCTCCAGATCTTTGGTATCTGCTGCGCCAGAACCTTGTGAGTGAC 900  
Db 602 ATAGGCATTGATTGCTGCGAGATATTTGGGATATGCTTGGCCCAAGATTTGGTTAGCGAT 661  
Qy 901 ATCAAGGCAGTGAAGGCCAACTGG 924  
Db 662 ATCGAAGCTGTCAGGCGGAGCTGG 685

RESULT 13

AAZ65259  
ID AAZ65259 standard; DNA; 2672 BP.

AC AAZ65259;

XX 23-MAR-2000 (first entry)

XX Human secreted protein gene 10.

XX Human; secreted protein; cancer; tumour; developmental abnormality;  
KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;  
therapy; chromosome 10; ds.

XX Homo sapiens.

OS WO9958660-A1.

XX 18-NOV-1999.

XX 06-MAY-1999; 99WO-US09847.

XX 12-MAY-1998; 98US-0085093.

PR 12-MAY-1998; 98US-0085094.

PR 12-MAY-1998; 98US-0085105.

PR 12-MAY-1998; 98US-0085180.

PR 18-MAY-1998; 98US-0085906.

PR 18-MAY-1998; 98US-0085920.

PR 18-MAY-1998; 98US-0085921.

PR 18-MAY-1998; 98US-0085922.

PR 18-MAY-1998; 98US-0085923.

PR 18-MAY-1998; 98US-0085924.

PR 18-MAY-1998; 98US-0085928.

PR 18-MAY-1998; 98US-0085925.

PR 18-MAY-1998; 98US-0085927.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Florence K, Ni J, Rosen CA, Carter KC, Moore PA;

PI Olsen HS, Shi Y, Young PE, Wei F, Brewer LA, Soppet DR;

PI Lafleur DW, Endress GA, Ebner R;

XX WPI; 2000-062296/05.

DR P-PSDB; AAY76133.

CC AAY76124 to AAY76223 represent the secreted proteins encoded by the 97  
CC human genes. The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions,  
CC e.g. by protein or gene therapy. Also pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new genes. Specific  
CC uses are described for each of the 97 genes, based on which tissues they  
CC are most highly expressed in, and include developing products for the  
CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
CC and foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
CC disorders, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The sequences shown in AAY76224 to AAY76424 represent fragments of the  
CC secreted proteins.  
XX

SQ Sequence 2672 BP; 548 A; 747 C; 708 G; 661 T; 8 other;

Query Match 14.1%; Score 357.4; DB 21; Length 2672;

Best Local Similarity 64.3%; Pred. No. 4e-76;

Matches 564; Conservative 3; Mismatches 304; Indels 6; Gaps 2;

Qy 48 CCCAGGGCGGCGCTGGAGGGCCGATCCCGGCCCGGCTCGGTTCCCGGGCGCGGGC 107

Db 99 CCCGGGGCTGCCAGGAATTTCCGAGCGGGCGCGCGCGCTGCGCGCGCGCGSGC 158

Qy 108 GCGTGTCT--CACCATGCCGGGCAAGACACAGCACTTCCAGGAACCCGAGGTGGCTGCTG 165

Db 159 GGATTTGCTTCTCAGAGATGCACTATATATAGTACTCTAACGCCAAGGTGACGTGCTG 218

Qy 166 CGGGAATACTCTCTGTTGGCTTCAACATTTGTTTCTGGGTGCTGGAGCCCTGTTCCCT 225

Db 219 GTACAAGTACTCTCTTTCAGCTACAACATCATCTTCTGRTTGGCTGGAGTTGTTCCCT 278

Qy 226 GGCCATCGGCTCTGGGCGCTGGGTGAGAGGGTGTCTCTCCAACTCTCGCGCTGAC 285

Db 279 TGGAGTCGGGCTGTGGGCATGGAGCGAAGGGGTGCTGTCCGACCTCAACAAAGTGAC 338

Qy 286 CGATCTGGGAGGCTCGACCCCTGTGGCTGTTGTAGTGGTGGAGCGCTCATGTCCGT 345

Db 339 CCGGATGCTATGGAATCGACCTGTGTGTGCTGCTGATGTTGATGTTTCAC 398

Qy 346 GCTGGGCTTTGCGGCTGATCGGGCTCTCCGGAGAACACTTTCTGCTCAAGTTTCT 405

Db 399 CTTGGGGTTTCCGCGCTGCTGCGGGGCTCTGCGGAGAAATATCTGCTGCTCAACTTTT 458

Qy 406 CTCAGTGTTCCTTGGGCTCATCTTCTTCTGGAGCTGGGAAACAGGATCTTGGCCTCGT 465

Db 459 CTGTGGCACCATGCTGCTCATCTTCTTCTGGAGCTGGGCTGTGGCGCTGCTGGCCTTCT 518

Qy 466 ATTCAGGAGCTGGATTTCGAGACGAGCTCAATTTCTTCAATTAACAACACCTCAAGGCTA 525

Db 519 GTTCCAGGACTGGGTGAGGGAACGGTTCCGGGAGTTCTTCGAGAGCAACATCAAGTCTTA 578

Qy 526 TCGGATGATATGACCTCCAGAACCTCATTTGCTTGTCTCAGGAATATTTGCTTGTCTG 585

Db 579 CCGGGACGATATCGATCTGCAAAACCTCTGACTCCCTTCAGAAAGCTAACCAAGTCTG 638

Qy 586 CGGAGCCCGAGGCGCTTAATGACTGGAACCTCAATATCTATTTCACTGCACTGACTTGAA 645

Db 639 TGGCGCATATGGCCCTGAAAGACTGGGACCTCAACGCTTACTTCAATTTGACGCGTGCAG 698

Qy 646 CCGGAGCCGAGAGCGCTGGGGGTGCTTCTCTGCTGTGTCAGGAGCCCTCG--A 701

Db 699 CTACAGCCGAGAGAGTGGGGTCCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 758

Qy 702 TGTCTCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGGAGCTGGAGCAGGG 761

Db 759 AGTTGTGAACACACAGTGTGGATATGATGTCAAGATTTCAGCTGAAGAGCAAGTGGGATGA 818

Qy 762 CTCCTACACACAAAGGCTGTGTGGCCAGTTTGAAGAGTGGCTGCTGAGCAACCTGAT 821

XX New isolated human genes and the secreted polypeptides they encode,  
XX useful for diagnosis and treatment of e.g. cancers, neurological  
XX disorders, immune diseases, inflammation or blood disorders -

XX Claim 1; Page 301-302; 475pp; English.

XX AAZ65250 to AAZ65350 represent 97 isolated human secreted protein genes.  
XX This sequence was found to be present on human chromosome 10.

Db 819 GTCCATCTTACAGAAAGGCTGCATCCAGGCGCTGGAAAGCTGCTCCCGGGAACATTTA 878  
Qy 822 CGTGGTGGCTGGGCTCTTTGGGGATCGCTCTCCAGATCTTTGGTATCTGCTGGC 881  
Db 879 CATTGGCTGGGCTCTTATCCGCCATCTGCTGTTGCAGATATTTGGCATCTCTCTGGC 938  
Qy 882 CCAGAACCTTGTGAGTGACATCAAGGCGAGTGAAGGCC 918  
Db 939 AAGGACCTGATCTCAGACATCGAGCGAGTGAAGGCC 975

## RESULT 14

ABX70953

ID ABX70953 standard; cDNA; 2465 BP.

XX AC ABX70953;

XX DT 05-MAR-2003 (first entry)

XX DE Novel human cDNA sequence #178.

XX Kw Human; gene; ss; nervous system disorder; peripheral neuropathy;  
Kw Huntington's disease; amyotrophic lateral sclerosis; haemophilia;  
Kw neurodegenerative disease; Parkinson's disease; Alzheimer's disease;  
Kw autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;  
Kw insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;  
Kw ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;  
Kw fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;  
Kw coagulation disorder; cancer; tumour; inflammatory disease;  
Kw septic shock; Crohn's disease; anaphylaxis; proliferation; chemotactic;  
Kw differentiation; stem cell growth factor; haematopoiesis; chemokinetic;  
Kw haemostatic; antiinflammatory; expressed sequence tag; EST.

XX OS Homo sapiens.

XX PN WO200281731-A2.

XX PD 17-OCT-2002.

XX PF 29-JAN-2002; 2002WO-US01222.

XX PR 30-JAN-2001; 2001US-0774528.

XX PA (HYSE-) HYSEQ INC.

XX PI (GOOD/) GOODRICH R W.

XX PI Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QH, Ren F;

XX PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;

XX PI WPI; 2003-058563/05.

XX PT Novel polypeptide useful for treating neurodegenerative diseases,  
PT myeloid or lymphoid cell disorders, bone disorders, mechanical and  
PT traumatic disorders, coagulation disorders, and inflammatory diseases

XX PS Claim 1; Page -; 612pp; English.

XX CC This invention relates to the cDNA sequences encoding an isolated  
CC novel human polypeptide. The protein encoded by the nucleic acid of  
CC the invention is useful for treating central and peripheral nervous  
CC system diseases (e.g. peripheral neuropathy, Huntington's disease,  
CC amyotrophic lateral sclerosis); neurodegenerative diseases (e.g.  
CC Parkinson's disease, Alzheimer's disease); autoimmune disease (e.g.  
CC systemic lupus erythematosus, rheumatoid arthritis, insulin-dependent  
CC diabetes mellitus); myeloid or lymphoid cell disorders (e.g. anaemia  
CC and thrombocytopaenia); wounds, ulcers, burns; bone disorders (e.g.  
CC osteoporosis, osteoarthritis); mechanical and traumatic disorders (e.g.  
CC stroke, head trauma); lung or liver fibrosis; reperfusion injury in  
CC various tissues; bacterial, viral or fungal infections; allergic  
CC conditions such as allergic rhinitis, asthma; coagulation disorders  
CC (e.g. haemophilia); cancer and tumours; and inflammatory diseases (e.g.

CC septic shock, Crohn's disease, anaphylaxis). The protein may be used to  
CC inhibit the growth, infection or function of infectious agents such as  
CC bacteria, fungi, viruses, or to effect bodily characteristics, such as  
CC biorhythms or circadian cycles of rhythms. The protein may also  
CC have proliferation/differentiation, stem cell growth factor,  
CC haematopoiesis regulation, immune stimulating or suppressing,  
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,  
CC and antiinflammatory activities. The cDNA sequences of the invention are  
CC useful for expressing recombinant protein for analysis. The present  
CC sequence represents a novel human cDNA sequence of the invention,  
CC this sequence is an expressed sequence tag (EST) and was identified  
CC using subtractive hybridisation.

XX SQ Sequence 2465 BP; 474 A; 718 C; 673 G; 600 T; 0 other;

Query Match 14.0%; Score 354.8; DB 25; Length 2465;

Best Local Similarity 64.5%; Pred. No. 1.7e-75;

Matches 564; Conservative 0; Mismatches 302; Indels 8; Gaps 2;

Qy 49 CCAGGGCGGCGCTGGAGGGCGATCCCGCGCGCTCGGTTCCGGGCGCGCGGGCG 108  
Db 43 CCGGGTCCGATTTCTGTCGGGCGCGCGCGCTCGCGCGCGCGCGGATCT 102  
Qy 109 GCTGCTCACCATGCGCGGCAAGCACCAGCACTTCCAGGAACCGAGTCCGCTGTCGCG 168  
Db 103 GCTTCTCA----GAAGATGCATATATAGATACTTAAGCCAAAGTACGTCTGTGTA 158  
Qy 169 GAAATACTTCTGTTGGCTTCAACATTTCTTGGGTGCTGGGAGCCCTGTTCTCTGGC 228  
Db 159 CAAGTACCTCTCTTTCAGCTACCAACATCATCTTCTGTTGGGTGGAGTGTCTTCTCT 218  
Qy 229 CATCGGCTCTGGGCTGGGCTGGGCTGGGCTGGTCTCTCAACATCTCTGCGCTGACCGA 288  
Db 219 AGTCGGCTGTGGGCAATGAGCGAAAGGGTGTGCTGCTCCGACCTCACCAGATGACCCG 278  
Qy 289 TCTGGAGGCGCTCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 348  
Db 279 GATGATGGAATCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338  
Qy 349 GGGCTTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 408  
Db 339 GGGGTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398  
Qy 409 AGTGTCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 468  
Db 399 TGCCACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 458  
Qy 469 CAAGGACTGGATTCGAGACCGAGCTCAATTTCTTCAATTAACAACAACGCTCAAGCCCTATCG 528  
Db 459 CCAGGACTGGGTCGAGGACCGGTTTCGGGAGTTCTTCGAGAGCAACATCAAGTCTTACCCG 518  
Qy 529 GGATGACATTTGACCTCCAGAACCTCATTTGCTCAGGAATATTTGCTTGTGCTGCGG 588  
Db 519 GGAGCATATCGATCTGCAAAACCTCATCGACTCCCTTCAGAAAGCTTAACAGTGTCTGG 578  
Qy 589 AGCCCGAGGGCTATGATGACTGGAACTCAATATCTATTTCAACTGCACGTGTAACCC 648  
Db 579 GCATATATGCGCTTGAAGACTGGGACCTCAACGCTCTACTTCAATTGTCAGCGGTGCCAGCTA 638  
Qy 649 GAGCGAGAGCGCTCGGGGTCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 704  
Db 639 CAGCCGAGAGAGAGTCCGGGTCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698  
Qy 705 CCTCAACACCCAGTGTGCTATGATGTCGGGCTCAAACTGGAGCTGGAGCAGCAGGGCTC 764  
Db 699 TGTGAACACACAGTGTGATATGATGTCAGGATTCAGCTGAGAGCAAGTGGGATGAGTCT 758  
Qy 765 CATACACACCAAGGCTGTGGGCGCATGTTGAGAGTGGCTGCGAGGACAACTGATCT 824  
Db 759 CATCTTACGAAAGGCTGCTCATCCAGGCGCTGAAAGTGGCTCCCGCGGCAACATTTACAT 818  
Qy 825 GGTGCTGGGCTCTTTGTTGGGCGATCGCTCTCTCCAGATCTTTGCTGATCTGCTGCGGCCA 884

Db 819 TGTGGCTGGGCTCTTCATCGCCATCTCGCTGTGGCAGATATTGGCATCTTCTCTGCAAG 878

QY 885 GAACCTTGTGAGTGACATCAAGCAGTGAAGGCC 918

Db 879 GACGCTGATCTCAGACATCGAGGAGTGAAGGCC 912

RESULT 15

AAF90629

ID AAF90629 standard; cDNA; 813 BP.

XX

AC AAF90629;

XX

DT 04-MAY-2001 (first entry)

XX

DE Human TANGO 339 cDNA ORF, SEQ ID NO:2.

XX

XX Secreted protein; transmembrane protein; TANGO; human; drug screening;

KW activity modulator; expression modulator; cancer; immunological disorder;

KW cytostatic; immunomodulatory; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN WO200109162-A2.

XX

PD 08-FEB-2001.

XX

PF 31-JUL-2000; 2000WO-US20935.

XX

PR 30-JUL-1999; 99US-0365164.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Fraser CC, Sharp JD, Kirst SJ, Barnes TM, Wrighton N, Myers PS;

PI Pan Y;

XX

XX WPI; 2001-138647/14.

DR P-PSDB; AAB87034, AAB87035, AAB87036.

XX

XX Nucleic acids encoding secreted TANGO and MANGO polypeptides, useful

PT for the prevention, diagnosis and treatment of, e.g. cancers and immune

PT disorders -

XX

XX Claim 2; Page 220; 332pp; English.

XX

CC The invention relates to novel secreted/transmembrane proteins, and

CC nucleic acids encoding them. The novel proteins are designated TANGO 339,

CC TANGO 353, TANGO 358, TANGO 365, TANGO 369, TANGO 383, TANGO

CC 393, TANGO 402, MANGO 346 and MANGO 349 and are of human origin, and a

CC murine TANGO 393 is also included within the scope of the invention. The

CC invention also encompasses fragments and variants of the proteins of the

CC invention, and nucleic acids encoding them. The invention additionally

CC relates to host cells comprising a nucleic acid of the invention; methods

CC for the production of a protein of the invention; an antibody specific

CC for a protein of the invention; methods for detecting a protein or

CC nucleic acid of the invention; and methods of identifying agents which

CC bind to or modulate the activity of a protein of the invention. The novel

CC secreted proteins, nucleic acids encoding them, and antibodies against

CC them may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate expression or activity of the secreted

CC proteins. The secreted proteins of the invention may also be used to

CC identify modulators of expression or activity, which may be useful in

CC the treatment of disorders associated with the proteins of the

CC invention e.g., cancers and immunological disorders. The present

CC sequence represents cDNA encoding human TANGO protein of the invention.

XX

XX Sequence 813 BP; 163 A; 213 C; 235 G; 202 T; 0 other;

SQ

Query Match 13.9%; Score 353.4; DB 22; Length 813;

Best Local Similarity 67.0%; Pred. No. 2.3e-75;

Matches 518; Conservative 0; Mismatches 251; Indels 4; Gaps 1;

QY 150 CCGAGGTGGTCTCGGGGAAATCTTCTGTGGCTTCAACATTTGTTTCTGGGTGC 209

Db 26 CCAAGTCAAGTCTGCTGTTTCAAGTACCTCTTTTCAAGTACACATCATCTTCTGGTTGG 85

QY 210 TGGAGCCCTGTTTCTTGGCCATCGGCCCTTGGGCTTGGAGGAGGTTTCTCTCCA 269

Db 86 CTGGAGTTGTTCTTCTTGGAGTCGGGCTGTGGGATGGAGCGAAAGGTTGCTGTCCG 145

QY 270 ACATCTCGGCTGACCGGATCTGGAGGCTCGACCTGTGTGGCTGTGTAGTGGTTG 329

Db 146 ACCTCAACAAAGTGACCCCGGATGCAATGGAATCGACCTGTGTGGTGGTCTCTATGGTGG 205

QY 330 GAGGGCTCATGTCCGTGGGCTTTGGCGGCTGCACTCGGGGCTCTCCGGAGAACACTT 389

Db 206 GGTGTGTATGTTTCACTTGGGTTCCCGGCTGCGGTGGGGCTCTGGGAGAAATATCT 265

QY 390 TCTGCTCAAGTTTTTCTCAGTGTTCCTTGGGCTCATCTTCTTCTGGAGCTGGCAACAG 449

Db 266 GCTTGTCTCAACTTTTCTGTGGCACCATCGTCTCATCTTCTTCTGGAGCTGGCTGTGG 325

QY 450 GGATCTTGGCCTTGTATTTCAAGGACTGTATTCAGAGACCAGCTCAATTTCTTCAATACA 509

Db 326 CCGTGTGGCTTCTTCTTCCAGGACTGGGTGAGGAGCCGGTTCCGGGAGTTTCTTCGAGA 385

QY 510 ACAAGCTCAAGGCTATCGGGATGACATTGACCTCCAGAACCTCATTTGACTTTGCTCAGG 569

Db 386 GCACATCAAGTCTTACCGGAGCATATCGATCTGCAGAACTCATCTCGACTCCCTTCAGA 445

QY 570 AATATTGCTTGTCTGCGGAGCCCGAGGCTTAATGACTGGAACCTCAATATCTATTTCA 629

Db 446 AAGCTAACCCAGTGTGTGGCGCATATGGCCCTGAGAGACTGGGACCTCAAGCTCTACTTCA 505

QY 630 ACTGACTGACTGTGAACCCGAGCGAGAGCGCTGCGGGTGGCTTCTCTGTGTGTGTCA 689

Db 506 ATTGAGCGGTGCCAGCTACAGCCGAGAGAGTGGCGGGTCCCTTCTCTCTGTGGTGC 565

QY 690 GGGACCTCGC---ATGCTCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGG 745

Db 566 CAGATCTCGCGCAAAAAGTTGAAACACACAGTGTGGATATGATGTACAGATTACAGTGA 625

QY 746 AGCTGGAGCAGCGGCTCCATACACACAAAGGCTGTGTGGGCCAGTTTGAAGAGTGGC 805

Db 626 AGAGCAAGTGGGATGAGTCCATCTTACGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGC 685

QY 806 TCGAGGACAACTGATCGTGGTGGCTTGTGTGGGCTCGCTCTCTCTCCAGATCT 865

Db 686 TCCCGGGAACATTTACATTTGTGGTGGCTTCTTATCGCCATCTCGCTGTGTGAGATAT 745

QY 866 TTGGTATCTGCTGGCCCAAGACCTTGTGAGTGACATCAAGGAGTGAAGGCC 918

Db 746 TTGGCATCTTCTTGGCAAGGAGCTGTATCTCAGACATCGAGGAGTGAAGGCC 798

Search completed: November 21, 2003, 17:31:07  
Job time : 658 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2003, 17:13:09 ; Search time 9165 Seconds  
(without alignments)  
11328.820 Million cell updates/sec

```

Title: US-09-972-970-2
Perfect score: 2538
Sequence: 1 ccacgcgtccggccgagcc.....aaaaaaaaaaaaaaaaaa 2538

```

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

GenEmbl.*
1:  gb_ba.*
2:  gb_ba.*
3:  gb_htg.*
4:  gb_in.*
5:  gb_om.*
6:  gb_ov.*
7:  gb_pat.*
8:  gb_ph.*
9:  gb_pl.*
10: gb_pr.*
11: gb_ro.*
12: gb_sts.*
13: gb_sy.*
14: gb_un.*
15: gb_vi.*
16: em_ba.*
17: em_fun.*
18: em_hum.*
19: em_in.*
20: em_mu.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_un.*
30: em_vi.*
31: em_htg_hum.*
32: em_htg_inv.*
33: em_htg_other.*
34: em_htg_mus.*
35: em_htg_pln.*
36: em_htg_rod.*
37: em_htg_mam.*
38: em_htg_wrt.*
39: em_sy.*
40: em_htg_hum.*
41: em_htg_mus.*
42: em_htg_other.*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	960.8	37.9	2426	9	BC010405	BC010405 Homo sapi
2	790.4	31.1	3184	6	AX420466	AX420466 Sequence
3	745.4	29.4	1516	10	BC010346	BC010346 Mus muscu
4	710.4	28.0	1813	6	AX420468	AX420468 Sequence
5	580.2	22.9	4445	9	AX024427	AX024427 Homo sapi
6	511	20.1	1567	6	AX064178	AX064178 Sequence
7	465.2	18.3	1685	5	BC041304	BC041304 Xenopus l
8	456.4	18.0	3175	10	AF121344	AF121344 Mus muscu
9	441.6	17.4	1405	9	AF065389	AF065389 Homo sapi
10	441.6	17.4	1416	9	BC009704	BC009704 Homo sapi
11	409.2	16.1	1408	9	AF053455	AF053455 Homo sapi
12	368.8	14.5	2428	8	BC025568	BC025568 Mus muscu
13	368.8	14.5	2498	10	BC024611	BC024611 Mus muscu
14	368.8	14.5	2500	10	BC026574	BC026574 Mus muscu
15	353.4	13.9	2553	9	AF3111903	AF3111903 Homo sapi
16	352	13.9	2556	9	HSM801611	AL136638 Homo sapi
17	351.8	13.9	1388	6	AX440923	AX440923 Sequence
18	333.6	13.1	1110	6	BD135990	BD135990 Secretary
19	305.8	12.0	132832	9	AC091934	AC091934 Homo sapi
20	300.2	11.8	2502	9	BC002920	BC002920 Homo sapi
21	295.8	11.7	864	6	AX4343015	AX4343015 Sequence
22	294.2	11.6	368	9	AF174603	AF174603 Homo sapi
23	244.6	9.6	394	6	AX072723	AX072723 Sequence
24	234	9.2	2073	9	BC044244	BC044244 Homo sapi
25	233.2	9.2	852	6	AX247836	AX247836 Sequence
26	233	9.2	389	6	AX072711	AX072711 Sequence
27	223.6	8.8	1995	10	BC024685	BC024685 Mus muscu
28	175	6.9	1998	3	AK116798	AK116798 Ciona int
29	165	6.5	1216	3	AF274013	AF274013 Drosophill
30	164.2	6.5	168318	2	AC024042	AC024042 Homo sapi
31	160	6.3	187478	2	AC123758	AC123758 Mus muscu
32	160	6.3	215066	2	AC136719	AC136719 Mus muscu
33	152.8	6.0	102298	10	AL671671	AL671671 Mouse DNA
34	152.8	6.0	17419	2	AC135142	AC135142 Rattus no
35	152.8	6.0	234117	2	AC130985	AC130985 Rattus no
36	152.4	6.0	168318	2	AC024042	AC024042 Homo sapi
37	151.4	6.0	267	6	BD072489	BD072489 Secreted
38	149.2	5.9	123377	9	AC108210	AC108210 Homo sapi
39	149.2	5.9	158785	2	AC027699	AC027699 Homo sapi
40	147.6	5.8	1816	3	BT004914	BT004914 Drosophill
41	135.8	5.4	171419	2	AC135142	AC135142 Rattus no
42	131.8	5.2	1661	10	BC003872	BC003872 Mus muscu
43	127.8	5.0	1694	6	BD132536	BD132536 Secreted
44	127.8	5.0	1703	9	AF120266	AF120266 Homo sapi
45	127.8	5.0	1726	9	BC003157	BC003157 Homo sapi

## ALIGNMENTS

RESULT 1	BC010405	2426 bp	mRNA	linear	PRI 10-JUN-2003
LOCUS	BC010405				
DEFINITION	Homo sapiens hypothetical protein MGC14859, mRNA (cDNA clone MGC:14859 IMAGE:3621871), complete cds.				
ACCESSION	BC010405				
VERSION	BC010405.1	GI:14714540			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Eureleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 2426)				
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S.,Wagner,L., Shenmen,C.M., Schuler,G.D.,				



```

1079 TC-TTGGCCCTGGGTAGTACATACGATGAGCAACCTTTAAACCTTGGCATATTTTCATGTA 1137
1080 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1081 969 TCCTTGGCCCTGGGTAGTATACCTGAGCGCACTTTTAAATTTGATGATTTTCAATTA 1028
1082 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1083 1138 AAGATCCAGATATCCAGCATCTTGTGTAAGATGCGCATCCGGCCACAGCGCTCTTCTAT 1197
1084 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1085 1029 AAGATCCAGATATCCAGCATCTTGTGTAAGATGCGCATCCGGCCACAGCGCTCTTCTAT 1088
1086 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1087 1198 GGCCTTGTCTCTGGGATGTCGCTTCTCTCTGAGGAGCCACCTTCAACCCGTCG 1257
1088 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1089 1089 GGAACAATCTACCTTGGGACATGCTCTCCCAAGTCAACAAAGGCGCCAGCTGGCCGTTCTA 1148
1090 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1091 1258 C-----TGCTTGCTGACCTGAGGCTGGGAGCTGGGCTGCTCTCCACCTCTGCAAG 1308
1092 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1093 1149 CTCACCTAAGTGGCGCTGACCTTGTACACTAGGAGCTGGCTC--CGACCTCTGCAGG 1206
1094 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1095 1309 TTTTCCCTGCAATGCTCAGGCTGCTGTGTAAGATGCGCATCCGGCCACAGCGCTTGGAGC 1368
1096 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1097 1207 GTTATTTCTTGCA----CTTCGAGGCGCTGCGGCGCAATCTGAGTGAACAACGCGGAC 1262
1098 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1099 1369 GTGAAGAATTTGGGAGGCTGAGGCTGCCCAAGAGGCCACAGCCCTGGGAAGGCTTGG 1428
1100 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1101 1263 CTGAAGA-TGAGAGGCTGGACCCCGCTTTGAAGAGGCTGAGCCCTGGGAAGGCG--GG 1319
1102 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1103 1429 CCCTCTGGGGGCAAGATGGGTGCCACACCGTGGCCAGGAGTGGCCGAGGCTGGGATG 1488
1104 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1105 1320 CCTTGTGGGACTGCGGTGGGAGT-AGAGTGGCCAGGAGGCTCTGAGGGGTGGGATG 1378
1106 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1107 1489 GAGATCAGGAGGTTTGGGACGAGCTAGCTGGAGGCTGAGCTTGTACCCCATGGGA 1548
1108 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1109 1379 GGGGTGAGGACAAATTTGCAAAAGAGTAGCTGGAAGCC-----ATGGGA 1423
1110 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1111 1549 TGGGGAGAGCCCTTTTGGGGGCTGATGGTAGGACTCAGCTCTGTGTGGAACCTAGT 1608
1112 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1113 1424 CTGGCGGAGCCCTGTTTGGGGATCTGGATGGTGTGACTCC-----TAGGAGTCAAGT 1475
1114 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1115 1609 TCAAAATCTTCCAGTGGCTGTAGAGTGTGCTCTGACCACTAGAGGGCGGCCACAC 1668
1116 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1117 1476 TCAGCATCTTGGCGTGG-CTGCAGAGCTGCTGTGAGGCTAGAGGCGACCGCAGCCC 1534
1118 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1119 1669 AGCATACCTGGTCTGCTTCTTCTAGGACCAACCCACCGAGTACAGCCCTGTGCTGGT 1728
1120 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1121 1535 CACACTCCCTGGGCTGTGGCTTCTCCCGCAACCTCACTCTAGTAGAGCCTGTG----- 1587
1122 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1123 1729 GTGTCCACCTGCTTACTAGTCTTCTTGGGTTCATGTAATTTACAGCTTCTTAAAGAGC 1788
1124 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1125 1588 -----CTGCTACTAGCGCTCTGGGTTTGGAGAGTTTGGGAATTTCTCA--GAGC 1637
1126 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1127 1789 AGAGTGGCTCAGATTGGGGAAGCTGGCAGCTGTTCTCAGATCTGCACAAAGCGGT---- 1844
1128 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1129 1638 CAACTGGCTCAGGCTTGGGAAGCTGGCTGCTGCTGCTCAGCTCAGCTCATCAGCTATGT 1697
1130 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1131 1845 ----GTGTGTGAGTATTTGTGTAATCAAGAGAGGTTTGGCTAGTGGCCAGTCTTTTA 1900
1132 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1133 1698 GAAGGGGTGTATGAGTGTATCTCGCGCCCTCCCTGGGCTGGTCCAGAGATCTCAA 1757
1134 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1135 1901 ACTTACATGCTCTCAGGCGGCTGGGTATATAA-----ATAAGTAGGCTTTAGCTG 1956
1136 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1137 1758 ACTCCGATCCCTCGGGGCAAGTATGTTGTGTAATGATGAACAGGCGCTTGGATTG 1817
1138 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1139 1957 TGAGGCTTTGGGACTTTAATTTTCCCACTATCTCTCGGAGATGGGACATAGAGACAT 2016
1140 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1141 1818 GGAGCGCTGCTCACTTTGA--CTTTCCCACTGTGCTGGAG-----CAAA 1861
1142 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1143 2017 TCCTTTGCTCAGAAATACCTTGCATGATGAGTCTGATCGCTAAGGGCAACTGGCCTT 2076
1144 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1145 1862 GACATCGTGTGATGAGAAAGTTTGCACAATCT-----AGTCGGTAACAGCCACTTTCCTT 1916
1146 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1147 2077 GAGTGACATCAAGGGGTGTGGGAGCTGTGGCAAAACACAGATTCCCACTCAAAATGCT 2136
1148 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1149 1917 GAGACCAAGAGAGTGGGTGGGAGTGGGGGGAGAGACGGGTCCCGCTCTGACAGTGGC 1976
1150 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

2137 GCCTGTC--CTTCGTTGGGGCTAAATGGCTGTACAGCGAGAATCTAGGTAGTCTGTCTTA 2194
2138 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2139 1977 CGCTGCCATATTCAGGTGTAGCTAATTCCTCTGGTGTGGGAATGCAG-----GCCTA 2028
2140 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2141 2195 ATGGGAAAGCTCTGGGAAGCCCAAGAGCTAGATTTTTCATGTGAACCTATCCCGAGTTT 2254
2142 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2143 2029 ATGACAGAAATCTGGGAAGCCAGAAATACAGATTTCT-ATGTGAGATGCTCTGATTTT 2087
2144 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2145 2255 AAGTTGTTGACGTAATGAGAAACCTCTTAAACCTCTATAGTCAAAAGGTGTGGGG 2314
2146 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2147 2088 TAAAGTTGTTGGCAGAAATTAATTCAGAAATCAAAATCTGCAGGCCAAACAAGGTGCAGGAC 2147
2148 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2149 2315 CCATTTTACACCTCCCCCAACCATAGTCCCTCAGGACAGTGCCTCCATGGGAGCCCTG 2374
2150 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2151 2148 CAGCTTTGGCCCATGCCC-CTGTAGTCCCTCTGGGACAGTCAACCTGGG-GTCTG 2205
2152 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2153 2375 GTGAGTCCACGGTTTCAGGAATGCTGGGAATGCTGCAGGTGGG--GGCTTGTGGCCAG 2432
2154 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2155 2206 GCTGCTCTGTCAATTGAGGATGCTGGGCACTGCTGCCGGGTGGCCAGGATGCGGCAATG 2265
2156 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2157 2433 CACCATCTGGCTCCAGGTGTG-----GGCTGGGCTG 2466
2158 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2159 2266 TGCCCAAGCAATGTGGCTCTTGGCCCGCTGGCCAGTGTCTGGGCCCTGCACAGGCGCT 2325
2160 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2161 2467 GGCTATGGGTGGCTCTACATGCTACAATAAATGGGCTCATGATAAATAAATAAATAA 2526
2162 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2163 2326 GGCTGTAGTGGTTGTACATGCTACAATAAATGCAGCTGGCAGCATTAATAAATAAATAA 2385
2164 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2165 2527 AAAAAAAAAA 2538
2166 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2167 2386 AAAAAAAAAA 2397
2168 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
AX420466
LOCUS
DEFINITION
Sequence 1 from Patent WO0216603.
AX420466
ACCESSION
AX420466.1
GI:21524614
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Leiby, K.R.
23228, a human tetraepanin family member and uses thereof
Patent: WO 0216603-A 1 28-FEB-2002;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
location
1..3184
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
168..980
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35489.1"
/db_xref="GI:21524615"
/translation="MPGKHQHPQEPVCGKGYFLPGFNVFVVLGALFLAIGLWANG
EKGVLNLSALTDLGLDPLVWLVVVGVMVSLGPGAGCIGALRENTFLKFSVFLGL
IFFLSLATGILAFVKDWIRDLNLFINNKAIRDDIDQLNLDIDFAQYWSGCCGARG
PFDNLNLYFNCTDLNPSRRCGVFPFSCVPDPAEDLVNTQCYDVLKLELEQQQFI
HTKCGVGFQFVWLNLIWVAGVFNGLALQIFGICLAQLVSDIKAVKRW"
BASE COUNT
675 a 883 c 945 g 681 t
ORIGIN
Query Match
31.1%; Score 790.4; DB 6; Length 3184;
Best Local Similarity
92.3%; Pred. No. 3.7e-167;
Matches 855; Conservative 0; Mismatches 66; Indels 5; Gaps 2;
QY 4 CCGCTCCCGCGCAGCCCGCGGCTAGGCCCGGCGGCTCTAGCCAGGCGGCCCGTG 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 52 CGCATGAAGCGGCGAGCCCGGCTAGGCCCCCGGCGGCTCTAGCCACAGGGCGGCCGCG 111  
Qy 64 GAGGGCCGATCCCGGC-CCCGGCTCCGGTTCCCGGGCGGCGGCGGCTCTCACCATGC 122  
Db 112 GGGCGCTGGGCTGGCTCCCGGCTCCCGGTTCCCGGCGGCGGCGGCTCACCATGC 171  
Qy 123 CGGCGAGCAGCAGCAGCTTCCAGAACCCGAGTCCGCTCGCTCGGGAATATCTTCTGT 182  
Db 172 CCGGCAAGCAGCAGCAGCTTCCAGAACCTGAGTCCGCTCGCTCGGGAATATCTTCTGT 231  
Qy 183 TTGGCTTCAACATTTGTTTCTCGGCTGCTGGAGCCCTGTTCTCGGCGCATCGGCTCTGGG 242  
Db 232 TTGGCTTCAACATTTGTTTCTCGGCTGCTGGAGCCCTGTTCTCGGCTATCGGCTCTGGG 291  
Qy 243 CTGGGCTGAGAGGGTGTCTCTCCAACTCTCTGCGCTGACCGATCTGGGAGGCTCG 302  
Db 292 CTGGGCTGAGAGGGGTTCTCTCGAACATCTCAGCGCTGACAGATCTGGGAGGCTTG 351  
Qy 303 ACCCTGTGCTGCTGTTTGTAGTGTGGAGCGTCATGCTCGTCAAGTTTCTCGGCT 362  
Db 352 ACCCGTGTGCTGTTTGTAGTGTGGAGCGTCATGCTCGTGGGCTTTGCTGGCT 411  
Qy 363 GCATC3GGGCTCTCCGGGAGAACACTTCTCTGCTCAAGTTTCTCAGTGTCTTCTGGCC 422  
Db 412 GCATGGGGCTCTCCGGGAGAACACTTCTCTGCTCAAGTTTCTCGTGTCTCTCGGTC 471  
Qy 423 TCATCTTCTCTGGAGCTGGCAACAGGATCTTGGGCTTCGTATTCAGGACTGGATTC 482  
Db 472 TCATCTTCTCTGGAGCTGGCAACAGGATCTTGGGCTTTGTCTTCAAGGACTGGATTC 531  
Qy 483 GAGACAGCTCAATTTCTTCAATCAACAAACAGTCAAGGCTATCGGGATGACATTAACC 542  
Db 532 GAGACAGCTCAACCTTCTTCAATCAACAAACAGTCAAGGCTATCGGGAGCAATTAACC 591  
Qy 543 TCCAGAACCTCAATGACTTTGCTCAGGAATATTGGTCTTGTGCGGAGCCCGAGGGCCTA 602  
Db 592 TCCAGAACCTCAATGACTTTGCTCAGGAATATTGGTCTTGTGAGCCCGAGGCCCA 651  
Qy 603 ATGACTGGAACTCAATATCTTCACTGCACTGATTAACCCGAGCCGAGAGGCT 662  
Db 652 ATGACTGGAACTCAATATCTTCACTGCACTGATTAACCCGAGCCGAGAGGCT 711  
Qy 663 CGCGGTGCTCTCTCTCTGCTGTGCTCAGGACCTCTG- - -GATGCTCTCAACACCCAGT 718  
Db 712 CGCGGTGCTCTCTCTCTGCTGTGCTCAGGACCTCTGCGAGGATGCTCTCAACCCAGT 771  
Qy 719 GTGCTATGATGTCCGCTCAAACTGGAGCTGAGCAGCAGGCTCCATACACCAAAAG 778  
Db 772 GTGCTACGACGTCCGCTCAAACTGGAGCTGAGCAGCAGGCTTCATCCACCAAAAG 831  
Qy 779 GCTGTGGGCGGAGTTGAGAGTGGCTGAGCAGCAGCAGTATCGTGGTGGCTGGGCTCT 838  
Db 832 GCTGTGGGCGGAGTTGAGAGTGGCTGAGCAGCAGCAGTATGTTGTGGCGGAGTCT 891  
Qy 839 TTGTGGGCGATCGCTCTCTCTCAGATCTTGTGATCTGCTGCGGCGGAGACCTTGTGAGTG 898  
Db 892 TCATGGGCGATCGGCTCTCTCAGATCTTGTGATCTGCTGCGGCGGAGACCTCTGAGTG 951  
Qy 899 ACATCAAGGCGAGTGAAGGCGCACTGG 924  
Db 952 ACATCAAGGCGAGTGAAGGCGCACTGG 977

RESULT 3  
BC010346 1516 bp mRNA linear ROD 16-APR-2003  
LOCUS Mus musculus RIKEN cDNA 2210021G21 gene, mRNA (cDNA clone MGC:6941  
IMAGE:2811935), complete cds.  
ACCESSION BC010346  
VERSION MGC.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1516)  
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Uedini, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
2238257  
12477932  
2 (bases 1 to 1516)  
Strausberg, R.  
Direct Submission  
Submitted (05-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [gcaps-r@mail.nih.gov](mailto:gcaps-r@mail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisse, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 5 Row: f Column: 18.

FEATURES  
source

1. .1516  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="MGC:6941 IMAGE:2811935"  
/tissue type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."  
/clone\_lib="NCI CGAP\_Mam5"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. .1516  
/gene="2210021G21Rik"  
/db\_xref="LocusID:74257"  
/db\_xref="MGI:1921507"  
136. .948  
/codon\_start=1  
/product="RIKEN cDNA 2210021G21"  
/protein\_id="AAH10346.1"  
/db\_xref="GI:16307593"  
/db\_xref="LocusID:74257"

## Gene

## CDS



/translation="MPGKHQFQDPEVCGCKYFLFGFNIVFWLGLALFLAIGLWGL  
EKGLNISALTDLGLDPLVFLVVGVMVSLGFRAGCIGALRENTPLKFRSFVFLGL  
IFLELAAGTILAUFKDMIRDLNPIFNNNKAYRDLDLQNLIDFRAOYFWSGCCARG  
PDNWLNIYFNCLDNPSPRRCGVFSCCVDPFAEDVINTQCCTDIRKLELQQGSI  
YTKGCVQFSEKIQDNLIVAGVLVGLIALLQIFGLCLAQNLVSDIKAVKANW"  
BASE COUNT 277 a 417 c 461 g 361 t

BASE COUNT	277 a	417 c	461 g	361 t	
ORIGIN					
Query Match	29.4%	Score 745.4;	DB 10;	Length 1516;	
Best Local Similarity	89.1%;	Pred. No. 5.2e-157;			
Matches 817;	Conservative 0;	Mismatches 96;	Indels 4;	Gaps 1;	
Qy	12	GCGCAGCGCCGGGCTAGGCCCGGGCGGCTCTAGCCCGAGGGCGGCCCTGAGAGGCGG	71		
Db	29	GCCGACGCCACCGGGCTAGGCCCGGGCGGCTCTCAGGCGCTGGGGGGCCCGCAGAGGGCTG	88		
Qy	72	ATCCGGCCCGGGCTCCGGTTCGGGGCGGGCGGGCGGCTGCTCACCATGCCGGGCAAGC	131		
Db	89	GGCCGGTCAACGGCTCGGGTTCGGGACGGCAGGTGGCGCTCACCATGCCGGGCAAGC	148		
Qy	132	ACCAGCACTTCCAGGAACCCGAGTCCGCTGCTGCGGGAATACTTCTGTTTGGCTTCA	191		
Db	149	ACCAGCAATTTCCAGGACCTTGAGTCCGCTGCTGCGGGAATACTTCTGTTTGGCTTCA	208		
Qy	192	ACATTGTTTCTGGGTCTGGAGCCCTGTTCTTGGCCATCTGGGCTCTGGGCTTGGGGTG	251		
Db	209	ACATTGTTTCTGGGTCTGGAGCCCTGTTCTTGGCCATTTGGCTCTGGGCTTGGGGTG	268		
Qy	252	AGAAGGGTGTCTCTCAACATCTCTGCGTGAACGATCTGGGAGGCTTCGACCTGTGTGT	311		
Db	269	AGAAGGGCGTTCCTCCAAACATCTCGGCGCTGACAGATCTGGGCGGTCTTTGACCCCGTGT	328		
Qy	312	GGCTGTTTGTAGTGGTGGAGGCGTCAATGCTGCTGGCTTGGCTTGGCGGCTGCATCGGG	371		
Db	329	GGCTGTTTGTGGTGGGGAGTCAATGCTGATGTTGGGCTTCGCGGCGCTGCATTGGGG	388		
Qy	372	CTCTCCGGGAGAACACTTTCCTGCTCAAGTTTTTCTCAGTGTTCCTTGGGCTCATCTTCT	431		
Db	389	CCCTCCGGGAAAAACACTTCTCTGCTCAAGTTCTTCTCTGTGTTCTCTGGGCTCATCTTCT	448		
Qy	432	TCTCGAGCTGGCAACAGGATCTTGGCTTCGTATTCAGGACTGGATTCGAGACGAC	491		
Db	449	TCTCGAGCTGGCGCGGGGATCTTGGCTTCGTCTTCAAGGATTCGATTCGAGACCAAC	508		
Qy	492	TCAATTTCTTCAATTAAACAAACGCTCAAGGCTTATCGGATGACATTCAGCTCCAGAAC	551		
Db	509	TTAATCTTTCATCAACAAACATGTCAAAGCTTACCGGGATGATCTCGACCTTTCAGAAC	568		
Qy	552	TCATTCAGCTTGTCTAGGAATATTGGTCTTGTGCGGAGCCCGAGGCGCTTAATGACTGGA	611		
Db	569	TTATCGACTTGTCTAGGAATATCTGGTCTTGTGCTGTGTCGCCGAGGGCCCAATGACTGGA	628		
Qy	612	ACCTCAATATCTATTTCAACTGCACTGACTTTGAACCCGAGCGCGAGAGCGCTGCGGGTGC	671		
Db	629	ACCTCAATATCTACTTCAACTGCACTGACTTTGAACCCAGCCGTGAGCGCTGCGGGTGC	688		
Qy	672	CCTTCTCTGCTGTGTGAGGACCCCTGC-----GATGTCTCAACACCCAGTGTGGCTATG	727		
Db	689	CCTTTTCTGCTGTGTGAGGACCCACGCGAAGATGTCTCAATACCCAGTGTGGCTATG	748		
Qy	728	ATGTCCGGCTCAAACTGGAGCTCGAGCAGCGGGCTCCATACACCAAAGGCTGTGTGG	787		
Db	749	ACATCCGACTCAAACTGGAGCTCGAGCAGCGGGCTCCATCTACACCAAAGGCTGTGTGG	808		
Qy	788	GCCAGTTTGAAGTGGCTGCGAGGACAACTGATCTGGTGGCTGGGGCTTTTGTGGGCA	847		
Db	809	GCCAGTTTGAAGTGGCTACAGAGCAAACTGATCTGGTGGCGCGGGCTTGTGGGGA	868		
Qy	848	TGGCTCTCTCCAGATCTTTGGTATCTGCTGGCCAGAACCTTGTGAGTGACATCAAGG	907		
Db	869	TGGCCCTCTCTCCAGATATTGGCTCTGCTTGGCCAGAACCTTGTGAGTGACATCAAGG	928		
Qy	908	CAGTGAAGGCCCAACTGG	924		

Db	929	CGGTGAAGCCCACTGG	945
RESULT 4			
LOCUS	AX420468		
DEFINITION	Sequence 3 from Patent WO0216603.		
ACCESSION	AX420468		
VERSION	AX420468.1	GI:21524616	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Leiby, K.R.		
AUTHORS	23228, a human tetraspanin family member and uses thereof		
TITLE	Patent: WO 0216603-A 3 28-FEB-2002;		
JOURNAL	Millennium Pharmaceuticals, Inc. (US)		
FEATURES	Location/Qualifiers		
source	1..813		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	151 a	233 c	235 g
ORIGIN		194 t	
Query Match	28.0%	Score	710.4
Best Local Similarity	93.2%	Pred. No.	4.1e-149
Matches	755	Conservative	0
Mismatches	51	Indels	4
Gaps	1		
Qy	119	ATCCCGGGCAGCACAGCACTTCCAGGAACCCGAGTTCGGCTCGCGGAAATACTTTC	178
Db	1	ATCCCGGGCAGCACAGCACTTTCAGGAACCTTCAGGAGCTCGCGGAAATACTTTC	60
Qy	179	CTGTTTGGCTTCAACATTTGTTCTGGGTCTCGGAGCCCTGTTCTTGGCCATCGGCTC	238
Db	61	CTGTTTGGCTTCAACATTTGTTCTGGGTCTCGGAGCCCTGTTCTTGGCTATCGGCTC	120
Qy	239	TGGGCTTGGGTGAGAGGGTGTCTCTCAACATCTCTGCGCTGACCGCATCTGGGAGGC	298
Db	121	TGGGCTTGGGTGAGAGGGCGTTCCTCGAATCTCAGCGCTGACAGATCTGGGAGGC	180
Qy	299	CTCGACCTGTGTGGCTGTTTGTAGTGGTGGAGGGTCAATGCTCGGCTTGGCTTGGC	358
Db	181	CTTGACCCGCTGTGGCTGTTTGTGTAGTGGAGGGTCAATGCTCGGCTTGGCTTGGCT	240
Qy	359	GGCTGCATCGGGCTCTCGGGAGAACACTTTCCTCTCAAGTTTCTCAGTGTTCCTT	418
Db	241	GGCTGCATTTGGGGCCCTCCGGGAGAACACTTTCCTCTCAAGTTTCTCAGTGTTCCTC	300
Qy	419	GGCTCATCTTCTTCTCGAGGTGGCAACAGGATCTTGGCCCTTCTGATTCAGGAGCTG	478
Db	301	GGTCTCATCTTCTTCTCGAGGTGGCAACAGGATCTTGGCCCTTGTCTTCAAGGAGCTG	360
Qy	479	ATTGAGACAGCTCAATTTCTTCAATTAACAAACAGTCAAGGCTTATCGGATGACATT	538
Db	361	ATTGAGACAGCTCAACCTTTCATCAACAAACAGTCAAGGCTTATCGGAGCAGATT	420
Qy	539	GACCTCCAGAACCTCAATGACTTTCCTCAGGAATATTTGCTTGGCTCGGAGCCGAGGG	598
Db	421	GACCTCCAGAACCTCAATGACTTTCCTCAGGAATATTTGCTTGGCTTGGCTCGGAGCCGAGGG	480
Qy	599	CCTAATGACTGGAACCTCAATATCTTATTTCAACTGCACTGACTTGAACCCGAGCCGAGAG	658
Db	481	CCCAATGACTGGAACCTCAATATCTTCAACTGCACTGACTTGAACCCGAGCCGAGAG	540
Qy	659	CGCTCGGGGTGCCCTTCTTCTGCTGTGTGAGGAGCCCTTGC---GATGTCTTCAACACC	714
Db	541	CGCTCGGGGTGCCCTTCTTCTGCTGTGTGAGGAGCCCTTGCCTCGGAGGATGTCTTCAACACC	600
Qy	715	CAGTGTGGCTATGATGTCCGGCTCAACTGGAGCTGGAGCAGGAGCTCCATACACACC	774

Db 601 CAGTGTGGCTACGACGTCGGCTCAACTGAGCTGGAGCAGAGGGCTTCAATCCACACC 660  
Qy 775 AAAGGCTGTGTGGCCAGTTTGAAGAAGTGGCTGCAGGACACCTGATGCTGGTGGCTGGG 834  
Db 661 AAAGGCTGCGTGGCCAGTTTGAAGAAGTGGCTGCAGGACACCTGATGCTGGTGGCTGGG 720  
Qy 835 GTCTTTTGGGTCATGCTCTCTCCAGATCTTTCGATCTTTCGATCTTTCGATCTTTCGAT 894  
Db 721 GTCTTTTGGGTCATGCTCTCTCCAGATCTTTCGATCTTTCGATCTTTCGATCTTTCGAT 780  
Qy 895 AGTGACATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 924  
Db 781 AGTGACATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 810

RESULT 5  
AK024427 4445 bp mRNA linear PRI 29-SEP-2000  
LOCUS Homo sapiens mRNA for FLJ00016 protein, partial cds.  
DEFINITION AK024427  
ACCESSION AK024427  
VERSION AK024427.1 GI:10440362  
KEYWORDS fls (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 4445)  
AUTHORS Ohara.O., Nagase.T., Kikuno.R. and Okumura.K.  
TITLE The nucleotide sequence of a long cDNA clone isolated from human spleen  
JOURNAL Published Only in DataBase (2000)  
AUTHORS Ohara.O., Nagase.T., Kikuno.R. and Okumura.K.  
TITLE Direct Submission  
JOURNAL Submitted (24-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/NEDO, Tel:81-438-52-3913, Fax:81-438-52-3914)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- and 3'-end one pass sequencing and clone selection; Kazusa DNA Research Institute.  
FEATURES  
source Location/Qualifiers  
1. .4445  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="as0016"  
/issue\_type="spleen"  
/dev\_stage="adult"  
/notes="vector:pBluescriptII SK plus"  
1. .4445  
/gene="FLJ00016"  
/note="For this clone, GeneMark analysis triggered an alert for spurious CDS split. The result of GeneMark analysis of the nucleotide sequence of this clone is accessible through http://www.kazusa.or.jp/NEDO. Start codon is not identified."  
/codon\_start=1  
/product="FLJ00016 protein"  
/protein\_id="BAB15717.1"  
/db\_xref="GI:10440363"  
/translation="GVLGDDGVWDSDSRHGGNGVARLWQDPPVPALKRGLGRLT SCASVGLLQTMGPVSAATGLGPASQVWRKQSLNCPPTSLFLASIGSKG ELVSPSLAPLCHOPGWAQACGGEVATHLAVPALSTALYSCSLGLFPNVRLLGW AMLTWGSPLPROMSCGARGPNDNLNIYFNCTDLPNSRERGVPFSCVPRPAEDVLA NTQCQYDRLKLELEQQGFIHTKGCQVFEKWLQDNLIVVAGVFMGIALQIFGLCLA QNLVSDIKAVKANWSKWNDDFENHWTPTTISEVLSTAGPQQNSLTGAPGAPPSRHVF

gene  
CDS

BASE COUNT 890 a 1278 c 1321 g 956 t  
ORIGIN  
Query Match 22.9%; Score 580.2; DB 9; Length 4445;  
Best Local Similarity 67.0%; Pred. No. 8.5e-120;  
Matches 1257; Conservative 0; Mismatches 483; Indels 137; Gaps 25;  
Qy 571 ATATGTGCTTGTGCGAGCGGCGGCTTAATGACTGGAACCTCAATATCTATTTCAA 630  
Db 2553 ACAGTGTCTTGTGCGAGCGGCGGCTTAATGACTGGAACCTCAATATCTATTTCAA 2612  
Qy 631 CTGCACCTGACTTGAACCCGAGCGGCGGCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 690  
Db 2613 CTGCACCTGACTTGAACCCGAGCGGCGGCTTCCCTCTCTCTCTCTCTCTCTCTCTCT 2672  
Qy 691 GGACCCCTGCG---GATGTCTCAACACCCAGTGTGGCTATGATGTCGGGCTCAAACTGGA 746  
Db 2673 GGACCCCTGCGGAGGATGTCTCAACACCCAGTGTGGCTACGCTCGGCTCAAACTGGA 2732  
Qy 747 GCTGGAGCAGCAGGCTCCATACACCAAGGCTGTGTGGCGCAGTTTGAAGAGTGGCT 806  
Db 2733 GCTGGAGCAGCAGGCTTTCATCCACCAAGGCTGTGTGGCGCAGTTTGAAGAGTGGCT 2792  
Qy 807 GCAGGACCACTGATCTGTGTGGCTGGGCTTTGTGGGCGATCGCTCTCTCTCCAGATCTT 866  
Db 2793 GCAGGACCACTGATCTGTGTGGCGGAGTCTTCATGGGCGATCGCCCTCTCTCCAGATCTT 2852  
Qy 867 TGTGTATCTGCTGGCGCAGAACCTTGTGAGTGACATCAAGGCGAGTGAAGGCCAACTGGAT 926  
Db 2853 TGGCATCTGCTGGCGCAGAACCTTGTGAGTGACATCAAGGCGAGTGAAGGCCAACTGGAG 2912  
Qy 927 CAACATGATGATGGCTACAACTACTCAAAATAAACAACCTTTGAAAAACCACTGGCTTA 986  
Db 2913 CAATGGAATGATG-----ACTTTGAAAAACCACTGGCTTA 2947  
Qy 987 CGCCCACTCTCAGAGTTCCATGGCGCAGGCGCTCAGCGCTGCGCTGCTGGCTGGG 1046  
Db 2948 CGCCCACTCTCAGAGTTCCATGGCGCAGGCGCTCAGCGCTGCGCTGCTGGCTGGG 3007  
Qy 1047 CCCC-AGCC-CAGACCCACCTCGCAACATGTTTC-TTGGCTGGGTAGTACATACGATG 1104  
Db 3008 CCCC-TGGCGCGCGCCACCCAGCGACATGTTTCTTTGGCTGGGTGTTTATACCCCTG 3067  
Qy 1105 AGCCAACTTTTAAACTTGGCATATTTCAATGTAAGTCCAGATCCCAGCATCTTGTGA 1164  
Db 3068 AGCCAACTTTTAAACTTGGTAGATTTCAATAAAGTCCAGATCCCAGCATCTTCTTGA 3127  
Qy 1165 AGAATGGCCATCGGCGCAGCGGCTCTTCTATGGCTTCCTCTCTCTGGGATGCGCTTC 1224  
Db 3128 AGAATGACCACTTGGCTACCGCGGCTCTTCTGGTGGCAACACTACCTGGGACACTGCGCTCC 3187  
Qy 1225 CTGTCTCTGAGGAGCCACCTCAGCCGCTGTC-----TGCTGCTGACCCCTGG 1275  
Db 3188 CCAGTCAACAGGCGCGCGCTGCGCTTCTACTCACCTAAGTGGCGCTGACCTTGT 3247  
Qy 1276 AGCTGGAGCTGGCTCTCTCAGCTCTGCAAGTTTTCCTCTGCAAACTGTCGAAGCT 1335  
Db 3248 ACATAGGAGCTGGCTCT--CCACCTCTGCGAGGTTATTTCTCTGCA---CCTCGAGGCC 3301  
Qy 1336 GCTGTGGCGCAAGCCCGGATCGAAGCTCGAGCGTCAAGATTTGGGAGGCTGGAGCTG 1395  
Db 3302 GCTGTGGCGCAATCTGAGTGAACACCGGAGCTGAAGGA-TGGAGAGCTGGAGCCCG 3360  
Qy 1396 CCCCAGAGGCGCAGCTCGGAAGGCTCTGCGCCCTCTCTGGGCGCAAGATGGGTGCCA 1455  
Db 3361 CTTTGAAGAGGCTGACGCTGGGAAGGG--CGGCTTGTCTGGGACTGCGGTGGAG-TA 3417  
Qy 1456 CGGTGCCAGGAGAGTGGCGGAGGCTGGGATGGAATCAGGAAGTTTGGGCGCAGACG 1515  
Db 3418 GAGTGTCCAGGAGGCTGTGAGGCTGGGATGGGCTCAGGACAAATTTTGAAGAAAG 3477  
Qy 1516 TAGCTGAAGCTGAGCTTGTTCACCCATGGGATGGGAGAGCCCTGTTTGGAGGGCGCT 1575

```
Db 3478 TAGCTGGAAGCCA-----TGGGACTGGCGGAGCCTGTTGGGGATCTG 3522
Qy 1576 GATGGTAGGACTCAGCCTCTGTTGGAACTCAGTTCAAAATCTCCAGTGGCTGTAGAG 1635
Db 3523 GATGGTTGACTCC-----TAGGAGTCAAGTTTCAGATCTTCGCGCGTG-CTGCAGAG 3573
Qy 1636 TTGCTCTCTGACCACTAGAGGGCGGCCACACAGCATTTACCTGGGTCTGCTCTTCTAG 1695
Db 3574 CTGCTGATGGGCACTAGAGGGCAGCCGCCACACTCCCTGGGTCTGGCTCTCTCC 3633
Qy 1696 GACAAACCCACAGATACAGCCCTGTGCTGGTGTGTCACCTGTCTACTAGTCTTTTG 1755
Db 3634 GCAACCTCACTCTAGTAGAGCTGTG-----CCTGCTACTAGGCTCTG 3678
Qy 1756 GPTTCATGGAATTTACAAGCTTTCTAAGAGCAGAGTGGCTCAGATTGGGGAAGCTGG 1815
Db 3679 GGGTTCGGAGAGTTTGGGAATTTCTCA--GAGCCAACTGGCTCAGGCTTGGGAAGCTGG 3736
Qy 1816 CAGCTGTCTCAGATCTGCACAAAGCGGT-----GTGTGGAGATATTCTGAATCA 1867
Db 3737 CTGCTGCCCTCAGCTCCGCTCATCAGCTATGTGAAGGGGTGTGTAGGAGTATCTCTG 3796
Qy 1868 AAGGAGAGTTTGGCTAGTGCCAGTCTTTTAACTTAGATGCCCTCAGGGCGGGTGG 1927
Db 3797 CGCCCTCTCCCTGGGCTGTGTCAGAGATCTCAAACTCCGATGCCCTGGGGCCACGTATG 3856
Qy 1928 TTATAAAA---ATAAGTAGGCTTTGAGCTGTGAGGCTTTGGGACTTTAATTTTTC 1983
Db 3857 TTGTGTAATGATGAACAGGCCCTTGAGTTGGAGCCGTGCTTCACTTTGA--CTTTC 3914
Qy 1984 CACTATTCTCGAGATGGGACATAGAGACAATGCTTTGTGCTGAGAAATCTTGCATG 2043
Db 3915 CACTGTTGCTGGAGA-----CAAGACATCTGTATGAGAGAAATTCGCAC 3960
Qy 2044 ATTGAGTCTGAGTCGTAGGGCACTGGCCCTTGTAGTGACATCAAGGGTGTGGGACT 2103
Db 3961 AATC-----TAGTCGGTAACAGCCACTTCTTGAGACCAAGAGAGTGGCGTGGGATG 4015
Qy 2104 GTGGCAAAACACAGATTCCCACTGAAATTTGGTGGCTGTC--CTTCCGTTGGGCTAATG 2161
Db 4016 GGGGAGAGACCGGTCTCCGCTGTGACAGTGGCCGTGCCATNTCAGGTGTAGCTAAT 4075
Qy 2162 GTGTACAGCGAATGTAGTAGGTCTGTCTAATGGGAGAGTCTGGAGAAAGCCAAAGAA 2221
Db 4076 GCTCTGGTGGGAATGCAG-----GCCTAATGACAGAAATCTGGAGAGCCAGAAA 4127
Qy 2222 GCTAGATTTTTCATGTGAATATCCCGAGTTTAAAGTTTGTGCACTAATGAGAAAAAC 2281
Db 4128 TACAGATTTGT-ATGTGAGATGTCTGTATTTTAAAGTTGTTGGCAGAAATTAATTCAGA 4186
Qy 2282 CTCTTAAACCTGATAGTCAAAAGGTGTGGGGCCATCTTTCACACCTCCGCCCATAG 2341
Db 4187 AATCAATCTGAGGCCAAACAGGTGCAAGCCAGCTTTGGCCCCATGGCCC-CTGTAG 4245
Qy 2342 GTCCCTCAGGGACAGTGGCCCATGGAGCCCTGGTGAATCCAGGTTTCAAGGAATGCTGG 2401
Db 4246 GTCCCTCTGGGACAGTCAAGCTGGG-GTCCCTGGCTGCTCTGTCAATGAGGATGCTGG 4304
Qy 2402 AACTGCTGAGTGGGC 2418
Db 4305 CACTGCTGCCGGTGGC 4321
```

```
RESULT 6
AX061778
LOCUS AX061778 1567 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 2 from Patent WO0078948.
ACCESSION AX061778
VERSION AX061778.1 GI:12539860
KEYWORDS Rattus sp.
SOURCE Rattus sp.
ORGANISM Rattus sp.
```

```
REFERENCE
AUTHORS Andriamanandry, C. and Maitre, M.
TITLE Cloning, expression and characterisation of a cDNA coding for a rat brain gamma-hydroxybutyrate (ghb) receptor
JOURNAL Patent: WO 0078948-A 2 28-DEC-2000;
FEATURES
Location/Qualifiers
source 1..1567
/organism="Rattus sp."
/mol_type="genomic DNA"
/db_xref="taxon:10118"
CDS 22..1560
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC25029.1"
/db_xref="GI:12539861"
/tranlation="MRRAAGLPGRLHAAARRGLGRPARVPRRAAGRRSPSPASIS
NSRTVLGAAGNTFCLASTLSSGCEPSPASGPGVRAFFPTSGGGIQQGLDPV
WLFVLTGMSVLGAGCIGALRETFLLKPFVFLGLIFFLAAGILAFVFKDMIR
DQNLFINNNKAYRDDIDLONLIDFAQBYSCCGARGDNWNLNRTSTALTSPSR
ERCYVFFCVARTLARKTSQPCYCTTAQTELEBHKIHLHOSWMPFEKWLKPDGWPGLG
ATASPMGLAGHPNSTSRQPTKLPQVMACTLAHRAVVGASKDALPHSONGLWD
VCYDLSPLCEHGTQEAETAGLWNGCTGAGLFTTELCTVWGVHMCVCVCVCVR
VCLCLVRVRGKHVICALVSTPGVSTPLVRWMPFQSGKDGARRVGVHPASPSLLMDV
SUGGACCLRLPHLHHDLPEAWSSPQCHSLEMPRLISLSRLPLRTHWTQDGY
ERLN"
BASE COUNT 281 a 436 c 480 g 370 t
ORIGIN
Query Match 20.1%; Score 511; DB 6; Length 1567;
Best Local Similarity 85.9%; Pred. No. 3.4e-104;
Matches 602; Conservative 0; Mismatches 95; Indels 4; Gaps 3;
Qy 12 GCGCAGAGCCGCGGCTAGGCCCGCGGCTCTAGCCAGGCGCGCGCTGAGGCGCG 71
Db 28 GCGCAGAGCCGCGGCTAGGCCCGCGGCTCTAGCCAGGCGCGCGCTGAGGCGCG 87
Qy 72 ATCCGCGCGCGCGCTCGGTTTCCCGGCGCGCGCGCGCTGCTCACATGCCGGCAAGC 131
Db 88 GCGCGGTTCCCGCTCGGTTTCCAGGCGCGGAGTGGCGCTCACATGCCGGCAAGC 147
Qy 132 ACCAGCACTTCCAGGAACCCGAGGTGCGTGTGCGGGGAAATACTCTCTGTTGGCTCA 191
Db 148 ATCAGCAATTCCAGGACCCCTGAGTTGGCTGTGCGGGAAATACTCTCTGTTGGCTCA 207
Qy 192 ACATTTGTTTCTGCGGTGCTGGAGCCCTGTTCTCGGCCATCGGCTCTGGGCTGGGTG 251
Db 208 ACATTTGTTTCTGCGGTGCTGGAGCCCTGTTCTCGGCCATCGGCTCTGGGCTGGGTG 267
Qy 252 AGAAGGTGTTTCTCTCAACATCTCTGCGCTGACCGAT-CTGGGAGGCTCTGACCTGTG 310
Db 268 AGAAGGTGTTTCTTTCACATCTCTCAGGCGGACAGATCCAGGCGGTCTTGACCCGCTG 327
Qy 311 TGGCTGTTTGTAGTGTGGAGCGTCAATGTCCTGCTGCGGTGCTGCGGCTGCATCGGG 370
Db 328 TGGCTGTTTGTAGTGTGGGGAATCATGTCAATGCTGGCTTGGCTGCGGCTGCATCGG 387
Qy 371 GCTCTCCGGGAGAACACTTCTCTGCTCAAGTTTCTCAGTCTGCTGCTGCTCATCTTC 430
Db 388 GCCCTCCGGGAAACACCTTCTCTGCTCAAAATTTTCTCTGTTCTCTCGGCTCATCTTC 447
Qy 431 TTCTGGAGCTGGCAACAGGATCTTGGCTCTGTTTCAAGGACTTGGATTCAGACCAAG 490
Db 448 TTCTGGAGCTGGCGCGCGGATCTGCGCTCTGTTTCAAGGATTTGATTCGAGACCA 507
Qy 491 CTCAATTTCTTCAATTAAACAACTCAAGGCTATCGGATGACATTTGACCTCCAGAAC 550
Db 508 CTTAACCTCTTCAATCAACAACTCAAGGCTATCGGATGACATTTGACCTCCAGAAC 567
Qy 551 CTCATTGACTTTGCTCAGGAATATTGTTCTGCTCGGGAGCCCGGCTTAATGACTGG 610
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

568 CTTATCGACTTGTCTAGGAATACTGCTGTGGAGCCCGAGGCCCAATGATGG 627
611 AACTCAATATC--TATTTCAACTGCACTGACTT-GAACCGGAGCCGAGAGCGCTGCGGG 667
628 AACCTCAACATCCGAGCTTCAACTGCACTGACTTCAAAACCCAGCCGAGCGCTGTGGG 687
668 GTGCGCTTCTCTGCTGTGTCAGGAGCCCTGCGATGTCCTC 708
688 GTGCGCTTCTCTGCTGTGTAAGAACCTCGGGAAGACGTC 728

```

BC041304 1685 bp mRNA linear VRT 14-JAN-2003  
 Xenopus laevis. Similar to transmembrane 4 superfamily member 9,  
 clone IMAGE:4683897, mRNA, partial cds.

BC041304  
 BC041304.1 GI:27735442

Xenopus laevis (African clawed frog)

Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 Xenopodinae; Xenopus.

1 (bases 1 to 1685)

Klein, S. and Strausberg, R.

Submitted (16-DEC-2002) National Institutes of Health, Xenopus Gene  
 Collection (XGC), National Institute of Child Health and Human  
 Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD  
 20892-7510, USA

NIH-XGC Project

Contact: XGC help desk

Email: cgabps-remail.nih.gov

Tissue Procurement: Dr. Igor Dawid

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

http://www.systemsbio.org

contact: amadan@systemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 94 Row: 9 Column: 18  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Similarity but not  
 identity to protein.

#### FEATURES

```

Location/Qualifiers
1..1685
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone_type="Embryo, stage 31/32, Xenopus"
/clone_lib="NICHG_XGC_Emb4"
/lab_host="DH105"
/notes="Vector: pCMV-SPORT6"
<1..926
/codon_start=3
/product="Similar to transmembrane 4 superfamily member 9"
/protein_id="AAH41304.1"
/db_xref="GI:27735443"
/translation="FGSTRVGLIYCFHLAPRAVSPDPPRCQPNLPSCSGMSGKH
YKGEVSCCKIFIFGNVIFWLLGLTFLVGLVWSEKGVLSNLSISITDLGGFPVW
LFLVGVGVNFIILGPACIGALRENTFLKFFSVFLGIIIFLELTAGVLAFFKDWIKD
QLQFTNNIRAYRDDIDQLNIDFTQYVQCAGGADDMNLINLYFNCTDSNRSRER
CGVPFSCCKDPAEDVINTCCGVVDVROKPELQQTITHTKGCVPFKEWLQNLTIIVA
GVFIQALLIFGICLALQNLVSDIENVRASW"

```

#### CDS

```

456 a 332 c 383 g 514 t

```

#### BASE COUNT

#### ORIGIN

```

Query Match 18.3%; Score 465.2; DB 5; Length 1685;
Best Local Similarity 73.7%; Pred. No. 7.1e-94;
Matches 607; Conservative 0; Mismatches 213; Indels 4; Gaps 1;

QY 133 CCAGCACTTCCAGGAACCCGAGGTGCGCTGCTCGGGAAATATCTTCTGCTTTGGCTTCAA 192
DB 128 CAAGCACTCAAGGGACCCGAGGTGAGTCTGATCAAGTACTTCAATATTCGGCTTCAA 187
QY 193 CATTTGTTTCTGGGTGCTGGGAGCCCTGTTCTCTGGCCATCGGCTCTTGGGCTCGGGTGA 252
DB 188 TGTCACTTCTGGCTCTTGGGACTTAGCTTCTTGGAGTTGGTCTGTGGGCATGAGTGA 247
QY 253 GAAGGTGTTCTCTCAACATCTCTGGGTGACCGATCTCGGAGGCCCTGACCCCTGTGTG 312
DB 248 AAAGGTGTGCTTTTCCAAACATTTCTCTATCAAGATCTTGGGGGTTTGGATCCAGTGTG 307
QY 313 GCTGTTTGTAGTGTGGAGGCGTCAATGCTCGGTGCTGGGCTTTGCCGCTGCAATCGGGC 372
DB 308 GCTGTTTCTGGTAGTGGAGGGGTCAATGTTATCTTGGGCTTTGCAGGCTGTATGGGGC 367
QY 373 TCTCCGGAGAACACTTTCTCTCAAGTTTTTCTCAGTGTTCCTTGGCTCTCATTTCTT 432
DB 368 ACTCCGTGAGAACACCTTCTCTGTTGAAATTTTTTCTGCTGTTCTTGGAAATTTATTTCTT 427
QY 433 CTGGAGCTGGGCAACAGGGATCTTGGCCTTCTGTTATTCAGGACTGGATTTCGAGACCACT 492
DB 428 TCTGGAGCTCACGGCGGGAGTTCTTGCCTTCGTGTTTCAAGAGCTGGATTAAAGACCACT 487
QY 493 CAATTTCTTCATTAAACAAAGCTCAAGCGCTATCGGATGACATTTGACCTCCAGACCT 552
DB 488 GCAGTTCTTTTAAACAAACATCAGAGCTTACAGAGTGAATGACATGACATTCAGAACT 547
QY 553 CATTTGACTTTGCTCAGGAATATTGTTCTTCTGCTGGAGCCCGAGGGCTTAATGACTGAA 612
DB 548 TATTGACTTCAACACAGGAATATTGGCAGTGTCTGCGGGCGTTCGGAGCTGATGTTGAA 607
QY 613 CCTCAATATCTATTTCACTGACTGCTGTAACCCGAGCCGAGAGCGCTGCGGGGTGCC 672
DB 608 TTTAAACATTTATTTCACTGCAACCGACTCTAACCGGAGCCGAGAGAGATGTGGGGTGCC 667
QY 673 CTCTCTCTGCTGTGTGTCAGGAGCCCTGCTGCTGCTGCGGAGCCCGAGGGCTTAATGACTGAA 728
DB 668 ATTTCTCTGCTGCACAAAAGACCTGCTGAGGATGTAATAACACAGTGTGGCTATGA 727
QY 729 TGTCCGGCTCAAACTGGAGCTGGAGCAGCAGGCTCCATACACACCAAGGCTGTGTGGG 788
DB 728 TGTCAACAAAAGCTGAACTGATCAACAGGAGACAAATTCACAAAAGGCTGTGTCCC 787
QY 789 CCAGTTTGAAGAGTGTGCTCAGAGCAACCTGATCGTGTGGTGGCTGCGGGCTTTGTGGGCAT 848
DB 788 ACAGTTTGAAGAAATGGCTGCAGGATAACCTTACCATTGTGGCAGGGGTGTTCAATGGTAT 847
QY 849 CGCTCTCTCCAGATCTTTGGTATCTGCTGCGCCAGAACCTTGTGAGTGACATCAAGGC 908
DB 848 TGGGCTTTTACAGATATTGGAAATCTGTTAGCGCAGAAATTTGGTGAGCGCAATTTGAAGC 907
QY 909 AGTGAAGGCCAACTGGATCAAAACATGATGATGCTGCTCAAACTAC 952
DB 908 CGTCCGAGCCAGCTGTGTAGAAATTTATCTCTGATACAAAGTCC 951

```

#### RESULT 8

#### AF121344

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### AF121344

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

AF121344 3175 bp mRNA linear ROD 01-FEB-2000  
 Mus musculus tetraspanin Tspan-5 (Tspan5) mRNA, complete cds.

AF121344

AF121344

AF121344.1 GI:6841032

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3175)  
 AUTHORS Garcia-Frigola, C., de Lecea, L. and Soriano, E.  
 TITLE Mouse Tspan-5 cDNA cloning  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3175)  
 AUTHORS Garcia-Frigola, C., de Lecea, L. and Soriano, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JAN-1999) Dept. of Animal and Plant Cell Biology,  
 University of Barcelona, Av. Diagonal 645, Barcelona 08028, Spain  
 FEATURES  
 source 1..3175  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Swiss Webster/NIH"  
 /db\_xref="taxon:10090"  
 gene 1..3175  
 /gene="Tspan5"  
 CDS 388..1194  
 /gene="Tspan5"  
 /note="member of transmembrane 4 superfamily"  
 /codon\_start=1  
 /product="tetraspanin Tspan-5"  
 /protein\_id="AAP28869.1"  
 /db\_xref="GI:6841033"  
 /translations="MSGKHVKGPEVSCCIKYIFGRNVIFWFLGITFLGILWAWNEK  
 GVLSNISSITDLGGDFDPVWFLVGVGMFLGAGCIGALRENTFLKPFVFLGIIF  
 FLELTAGVLAFFVKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFIOEYWOCCAFGAD  
 DMLNLYFNCTDSNARERCVPFSCCTKDPADVINTOCGYDAROKPEVDQIVYI  
 KGCVPQFVKWLQNLITIVAGIFIGIALQLFGICLAQLVSDIEAVRASW"  
 BASE COUNT 667 a 811 c 840 g 857 t  
 ORIGIN  
 Query Match 18.0%; Score 456.4; DB 10; Length 3175;  
 Best Local Similarity 73.9%; Pred. No. 6.7e-92;  
 Matches 594; Conservative 0; Mismatches 206; Indels 4; Gaps 1;  
 QY 135 AGCACTTCAGGAACCGAGGTGGCTGCTGGCGGAAATACTCTCTGTTTGGCTTCAACA 194  
 DB 398 AGCACTACAGGGCTGCTGAAGTCAGTTGTCATCAATACTCTCTGTTTGGCTTCAATG 457  
 QY 195 TTGTTTCTGGGTCTGGGAGCCCTGTTCTGCGCCATCGCCTCTGGCGCTGGGTGAGA 254  
 DB 458 TCATATTTGGTTTTGGGAATACGTTTCTTGGAAATCGGACTGTGGCGGTGGAATGAA 517  
 QY 255 AGGGTGTCTCTCCAAACATCTCTGCGCTGACCATCTGGAGGCTCGACCTCTGTGGC 314  
 DB 518 AAGGTGTCTCTCCAAACATCTCTGCTCATCACCGACCTCGGTGGCTTGACCCAGTGGC 577  
 QY 315 TGTGTGTAGTGTGGAGGCTCATGTCGCTGCTGGCTTTGCCGCTGCATCGGGCTC 374  
 DB 578 TTTCCTGTGGTGGGAGGTGATGTTTCTTCTGGGGTTTGACGGGTGCATCGGAGCAC 637  
 QY 375 TCGGGAGAACACTTTCTCTCAAGTTTCTCAGTGTCTGCTGCTGCTGCTGCTGCTGCT 434  
 DB 638 TTGGGAAACACTTTCTCTCAAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 697  
 QY 435 TGAGCTGGCAACAGGATCTTGGCTCTGATTTCAAGACTGGATTCGAGACCACTCA 494  
 DB 698 TGGAACTCACTGCGGGTGTGGCATTTGTTTCAAGACTGGATTCGAGACCACTGT 757  
 QY 495 ATTTCTTCAATCAACACAGCTCAAGGCTATCGGATGACATTTGACCTCCAGAACCTCA 554  
 DB 758 ATTTCTTCAATCAACACAGCTCAAGGCTATCGGATGACATTTGATCTACAGAACCTCA 817  
 QY 555 TTGACTTTCTCAGGAATATTGCTCTGTGCTGGGAGCCCGAGGCTTATGACTGAACC 614  
 DB 818 TAGACTTCAACAGGATATTGGCAGTGTGTGGGCTTTTGAGCTGATGATGGAACC 877  
 QY 615 TCAATATCTATTCAACTGACTGACTTGAACCCGAGCGAGAGCGCTCGCGGGTGCCT 674  
 DB 878 TAAATATTACTTCAATTCACAGATTCCAAATGCAAGCGAGCGAGTGGGTGTGCCAT 937  
 QY 675 TCTCTGCTGTCTAGGAGCCCTGC-----GATGTCCTCAACACCCAGCTGTGGCTATGATG 730

DB 938 TTTCTGCTGCACCTAAAGACCCCGGGAAGATGTTCATCAACACTCAGTGTGGCTATGATG 997  
 QY 731 TCGGCTCAAACTGGAGCTGGAGCAGGCTCCATACACCAAGGCTGTGTGGGCC 790  
 DB 998 CCAGGCAGAAACAGAGATTGACCAACAGATTGTAATCTACACAAAAGGCTGTGTGCC 1057  
 QY 791 AGTTTGAAGAGTGGCTGCAGGACCACTGATCGTGGTGGCTGGGCTCTTTGTGGGCATCG 850  
 DB 1058 AGTTTGAAGAGTGGCTGCAGGACCAATTTAACCATCGTGGCTGGTATTTTCATAGCATG 1117  
 QY 851 CTCTCTCCAGATCTTTGGTATCTGCTGCCGCCAGAACCTTTGTGAGTGACATCAAGGCAG 910  
 DB 1118 CATTTGCTACAGATTTTGGGATATGCTGCGCAGAGATTTGGTGAGTGACATCAAGCTG 1177  
 QY 911 TGAAGGCCAACTGGATCAAAACATG 934  
 DB 1178 TCAGGGCTAGCTGGTAGACCCCTG 1201  
 RESULT 9  
 AF065389  
 LOCUS Homo sapiens tetraspan NET-4 mRNA, complete cds. PRI 28-APR-2000  
 DEFINITION AF065389  
 ACCESSION AF065389  
 VERSION AF065389.1 GI:3152702  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1405)  
 AUTHORS Serru, V., Dessen, P., Boucheix, C. and Rubinstein, E.  
 TITLE Sequence and expression of seven new tetraspans  
 JOURNAL Biochim. Biophys. Acta 1478 (1), 159-163 (2000)  
 MEDLINE 20185353  
 PUBMED 10719184  
 REFERENCE 2 (bases 1 to 1405)  
 AUTHORS Rubinstein, E., Serru, V. and Boucheix, C.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAY-1998) INSERM U268, 14 av Paul Vaillant Couturier,  
 Villejuif 94807, France  
 FEATURES  
 Location/Qualifiers  
 1..1405  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE ID 219547"  
 348..1154  
 /note="similar to Tspan-5; TM4SF"  
 /codon\_start=1  
 /product="tetraspan NET-4"  
 /protein\_id="AAC17120.1"  
 /db\_xref="GI:3152703"  
 /translation="MSGKHVKGPEVSCCIKYIFGRNVIFWFLGITFLGILWAWNEK  
 GVLSNISSITDLGGDFDPVWFLVGVGMFLGAGCIGALRENTFLKPFVFLGIIF  
 FLELTAGVLAFFVKDWIKDQLYFFINNNIRAYRDDIDLQNLIDFIOEYWOCCAFGAD  
 DMLNLYFNCTDSNARERCVPFSCCTKDPADVINTOCGYDAROKPEVDQIVYI  
 KGCVPQFVKWLQNLITIVAGIFIGIALQLFGICLAQLVSDIEAVRASW"  
 BASE COUNT 309 a 357 c 387 g 352 t  
 ORIGIN  
 Query Match 17.4%; Score 441.6; DB 9; Length 1405;  
 Best Local Similarity 73.2%; Pred. No. 1.5e-88;  
 Matches 581; Conservative 0; Mismatches 209; Indels 4; Gaps 1;  
 QY 135 AGCACTTCAGGAACCGAGGTGGCTGCTGGCGGAAATACTCTCTGTTTGGCTTCAACA 194  
 DB 358 AGCACTACAGGGCTGCTGAAGTCAGTTGTCATCAATACTCTATATTGCTTCAATG 417  
 QY 195 TTGTTTCTGGGTGTGGGAGCCCTGTTCTCGGCCATCGGCTCTGGGCTCGGGGTGAGA 254  
 DB 418 TCATATTTGGTTTTTGGGAATAACATTTCTTGGAAATGGACTGTGGGCATGGAATGAA 477

```
Qy 255 AGGTGTTCTCTCAACATCTCTGGCTGACCGATCTGGGAGGCTCGACCTCTGTGGC 314
Db 478 AAGGAGTTCTGTCCAAATCTCTTCCATCCGATCTCGGCGCTTTGACCCAGTTGGC 537
Qy 315 TGTTGTAGTGTGGAGGCTCATGTCGCTGGCTTTGGCGGCTGATCGGGGCTC 374
Db 538 TCCTCTGTGTGGAGGAGTGATGTTCAATTTGGGATTTGACGGTGATTTGGAGGCG 597
Qy 375 TCCGGGAGAACATCTTCTGCTCAAGTTTCTCAGTGTCTTCTTGGGCTCATCTTCTTC 434
Db 598 TACGGGAAACACTTCTCTCAAGTTTCTTCTGTTCTCTGGGAATATTTTCTTCTTC 657
Qy 435 TGGAGCTGGCAACAGGATCTTGGCTTCTGATTTCAAGGACTGATTCGAGACCACTCA 494
Db 658 TGGAGCTCACTGCCGAGTCTAGCATTTGTTTCAAGAGCTGGATCAAGAGACCACTGT 717
Qy 495 ATTTCTTCTTAAACAAACAGCTCAAGGCTATCGGGATGACATTTGACTTCAGAACCTCA 554
Db 718 ATTTCTTATAAACAACAACATCAGACATATCGGATGACATTTGATTTGCAAACTCA 777
Qy 555 TTGACTTTGCTCAGAAATATTTGCTTGTCTGGAGCGGAGCGGCTTAATGACTGGAACC 614
Db 778 TAGACTTCAACCCAGGAATATTTGGCAGTCTGTGGGCTTTTGGAGCTGATGATTTGGAACC 837
Qy 615 TCAATATCTATTTCAACTGCTGACTGACTGACCTGACCCGAGCGGAGCGGCTGGGGTGCCT 674
Db 838 TAAATATTTACTTCAATTTGACAGATTTCCAAATGCAAGTCGAGAGCGATGTGGCTCCAT 897
Qy 675 TCTCTCTGCTGTCTCAGGACCCCTGC---GATGTCTCTCAACACCCAGTGTGGCTATGATG 730
Db 898 TCTCTCTGCTGACTAAAGATCCCGCAGAGATGTCATCAACACTCAGTGTGGCTATGATG 957
Qy 731 TCCGGCTCAAACTGGAGCTGAGCAGAGGCTCCATACACACAAAGGCTGTGTGGGCC 790
Db 958 CCAGGCAAAACCAAGAGTTGACACGAGATTTGTAATCTACACAAAGGCTGTGTGCCCC 1017
Qy 791 AGTTTGAGAGTGTGCTGAGACAACTGATCGTGTGGCTGTGGGCTCTTTGTGGGCATCG 850
Db 1018 AGTTTGAGAGTGTGTCAGGACAAATTAACCATGTTGCTGTGATTTTATAGGCATG 1077
Qy 851 CTCTCTCCAGATCTTTGGTATCTGCTGCCCAGAACCTTTGTGAGTGACATCAAGGCAG 910
Db 1078 CATGTCTGCAGATATTTGGGATATGCTGCCCCAGAAATTTGTTAGCGATATCGAAGCTG 1137
Qy 911 TGAAGGCCAAGTGG 924
Db 1138 TCAGGGCGAGCTGG 1151
```

```
RESULT 10
LOCUS BC009704 1416 bp mRNA linear PRI 13-JAN-2003
DEFINITION Homo sapiens, tetraspan 5, clone MGC:9300 IMAGE:3895933, mRNA,
complete cds.
ACCESSION BC009704
VERSION BC009704.1 GI:16307230
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1416)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: ATCC
```

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdpaxil.stanford.edu](mailto:mdpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 14 Row: h Column: 4  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 21264582.

#### FEATURES

Location/Qualifiers  
1..1416  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:9300 IMAGE:3895933"  
/tissue\_type="Pancreas, epithelioid carcinoma"  
/clone\_lib="NIH MGC\_70"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
334..1140  
/codon\_start=1  
/product="tetraspan 5"  
/protein\_id="AAH09704.1"  
/db\_xref="GI:16307231"  
/db\_xref="LocusID:10098"  
/translations="MSGKHKGPEVSCCTKYIFGNVIFNPLGILFGLGLAWNEK  
GVLSNISTDLGGDFVWLFLVVGVMFLGAGCIGALRENTFLKRFVFLGILP  
FUELTAGVLAIPFKWDKQLYFFINNNRAYRDDIDLQNLIDFTQBYWCCQAGFAD  
DNLNLTFTNCTDSNARERCQFPFSCCTKDPADVDINTQCYDARQKPEVDQQIVYT  
KGCVPQFEKWLQNLTIAGIFIGIALLOIFGICLAQNLVSDIEAVRAS"  
BASE COUNT 325 a 354 c 385 g 352 t  
ORIGIN

Query Match 17.4%; Score 441.6; DB 9; Length 1416;  
Best Local Similarity 73.2%; Pred. No. 1.5e-88;  
Matches 581; Conservative 0; Mismatches 209; Indels 4; Gaps 1;  
Qy 135 AGCACTTCACGGAACCGAGGTCGGCTGCTGCGGGAATACTTCTCTGTGGCTTCACA 194  
Db 344 AGCACTACAGGTCCTGAAGTCAGTTGTTCATCAAAATCTTCATATTTGGCTTCAATG 403  
Qy 195 TTGTTTTCTGGGTGCTGGAGCCCTGTTCTGCGCCATCGGCCCTGCGGCTGGGGTGAGA 254  
Db 404 TCATATTTGGTTTTGGGAATAAATCTTCTTGAATTTGACTGTGGCATGGAATGAA 463  
Qy 255 AGGCTGTTCTCCAAACATCTTCGCGCTGACCATCTGGAGGCTCGACCTGTGTGGC 314  
Db 464 AAGGAGTTCTGTCCAAATCTCTTCCATCACCCTGTCGGCGCTTTGACCCAGTTTGGC 523  
Qy 315 TGTTTCTAGTGTGGAGGCTCATCTCGCTGCTGGCTTTGCCGCTGCTCGGGCTC 374  
Db 524 TCTTCTGTGGTGGAGGAGTGATGTTCAATTTGGGATTTGCAGGTTGCAATTTGGAGCGC 583  
Qy 375 TCCGGGAGAACACTTCTCTGCTCAAGTTTTTCTCAGTGTTCCTTGGGCTCATCTTCTTCC 434  
Db 584 TACGGGAAACACTTCTCTCTCAAGTTTTTCTGTTCTCTGGGAATATTTTCTTCTTC 643  
Qy 435 TGGAGCTGGCAACAGGATCTTGGCCCTGCTATATTCAGGACTGGATTCGAGACCACTCA 494  
Db 644 TGGAGCTCACTGCGCGAGTTCTAGCATTTGTTTCAAGAGACTGGATCAAGAGACCACTGT 703  
Qy 495 ATTTCTTCTTAAACAACAGCTCAAGGCTATCGGATGACATTTGACCTCCAGAACTCA 554  
Db 704 ATTTCTTATAAACAACAACATCAGAGCATATCGGATGACATTTGATTTTGAACAACTCA 763  
Qy 555 TTGACTTTGCTCAGGAATATTTGCTTCTGTGCGAGCCGCGGCTTAATGACTGGAACC 614  
Db 764 TAGACTTCAACCCAGGAATATTTGGCAGTCTGTGGGCTTTTGGAGCTGATGATTTGGAACC 823



Matches	570; Cons	570; Cons	0; Mismatches	208; Indels	16; Gaps	2;
Qy	135	AGCACTTCCAGGAACCCGAGGTGGCTGCTGCGGGAATACTTCTGTTTGGCTTCAACA	194			
Db	362	AGCACTACAGGGTCTGAAGTCAGTGTGTGTCATCAATACTTATATTTGGCTTCAATG	421			
Qy	195	TTGTTTTCTGGGTGCTGGGAGCCCTGTTCTCTGGCCATCGGCTCTGGGCTCGGGGTGACA	254			
Db	422	TCATATTTTGGTTTTTGGGAATAACATTTCTTGGAAATGGAGCTGTGGGCATGGAATGAAA	481			
Qy	255	AGGGTGTCTCTCCAACTCTCTGGCTGACCGATCTGGGAGGCGCTCGACCCCTGTGTGCG	314			
Db	482	AAGAGTTCTGTCCAACTCTCTTCATCACCAGATCTCGCGGCTTTGACCCAGTTTGGC	541			
Qy	315	TGTTGTAGTGGTTGGAGCGCTCATGTCGCTGCTGGCTTTTCCGCGCTGCATCGGGCTC	374			
Db	542	TCTTCTTGTGGTGGGAGAGTATGTTTCATTTTGGATTTGACGGTGCATTTGGAGCGC	601			
Qy	375	TCCGGGAGAACACTTTCCTGCTCAAGTTTTCCTCAGTGTTCTTGGCCTCATCTTCTTCC	434			
Db	602	TACGGGAAACACTTTTCC-----TTTCTGTGTCTCGGGAATATTTTCTTCTCC	649			
Qy	435	TGGAGCTGGCAACAGGGATCTTGGCTTCGTATTTCAAGGACTGGATTTCGAGACCAGCTCA	494			

Qy	495	ATTTCCTTCATTAAACAACACGTC	CAAGGCCCTATCGGATGACATTC	GACCTCCAGAACCTCA	554
Db	710	ATTTCTTTATAACAACAACATC	CAGGCATATCGGGATGACATTT	TGTTGCAAAACCTCA	769
Qy	555	TTGACTTTTGCTCAGGAATATTGGT	CTTGTCGGAGCCGAGGGCCTAAT	GTACTGGAACC	614
Db	770	TAGACTTTCACCCAGGAATATTGGC	AGTGCTGTGGGGCTTTTGGAGCT	GTATGATTGGAACC	829
Qy	615	TCAAATATCTATTTCACCTGCACT	GTAAACCGAGCGGAGAGCGCTC	CGGGGTGCCCT	674
Db	830	TAAATATTTTACTTTCAATTTGCA	CAGATTCCAATGCAAGTCGAGG	CGATGTGGCGTTCCAT	889
Qy	675	TCCTCTGCTGTGTCAGGGACCTGC	----GATGTCTCTCAACACCCAG	ATGTGCTATGATG	730
Db	890	TCCTCTGTGCTCACTAAAGATCCG	CAGAAGATGTCATCAACACTCAG	TGTGCTATGATG	949
Qy	731	TCGGCTTCAAACTGGAGCTGGAGC	AGCGGGCTCCATACACCAAAGCT	TGTGTGGGCC	790
Db	950	CCAGGCAAAAACAGAAAGTTGAC	CAGCAGATTGTAATCTTACACGA	AAAGCTGTGTGCCCC	1009
Qy	791	AGTTTTCAGAAAGTGGCTGCAAGG	ACAACCTTGATCGTGGTCTGGG	GTCTTTTGTGGGCATCG	850

[illegible]

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 2428)  
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,



Klausner R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T.I., Max, S.I., Wang, J., Heien, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, W.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2428)

Strausberg, R.

Direct Submission

Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapb-re@mail.nih.gov](mailto:cgapb-re@mail.nih.gov)

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NIHC)

Gaithersburg, Maryland

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc\\_mgc@hghri.nih.gov](mailto:nisc_mgc@hghri.nih.gov)

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 56 Row: j Column: 21

This clone was selected for full length sequencing because it passed the following selection criteria: GenomesScan gene prediction.

## FEATURES

## Location/Qualifiers

1..2428

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="MGC:36595 IMAGE:5322531"

/tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."

/clone\_lib="NCI CGAP\_Mam6"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..2428

/gene="D14Etd226e"

/note="synonym: MGC36554"

/db\_xref="LocusID:52588"

/db\_xref="MGI:1196325"

64..876

## gene

## CDS

/codon\_start=1

/product="tetraspanin similar to TM4SF9"

/protein\_id="AAH25568.1"

/db\_xref="GI:19343884"

/db\_xref="LocusID:52588"

/translation="MHYRYSNAEVSCKYLLFSLVFLAGVFLVGLWAMSEK  
GVLSDLTKVTRLHGIDPVVLVLMVGVMTFLGAGCVGALRENI<sup>1</sup>CLLKKFCCGAILVLF  
FLELAVAVLAF<sup>2</sup>PDQWVRFRFPFESN<sup>3</sup>IKSYR<sup>4</sup>DDLDLNLISL<sup>5</sup>OKANOC<sup>6</sup>CGAYPE  
DMLNLYFNC<sup>7</sup>SGASYR<sup>8</sup>CKGV<sup>9</sup>CCVFP<sup>10</sup>PAOKV<sup>11</sup>NTCGY<sup>12</sup>DR<sup>13</sup>OLK<sup>14</sup>SNDE<sup>15</sup>FI<sup>16</sup>FT  
KGCIOALEGWLPRNIYIVAGVFI<sup>17</sup>LSLLOL<sup>18</sup>FI<sup>19</sup>FLARTLIS<sup>20</sup>DI<sup>21</sup>EAVK<sup>22</sup>AGH<sup>23</sup>F"

BASE COUNT 517 a 640 c 624 g 647 t

ORIGIN

Query Match 14.5%; Score 368.8; DB 10; Length 2428;

Best Local Similarity 67.9%; Pred. No. 3.7e-72;

Matches 531; Conservative 0; Mismatches 247; Indels 4; Gaps 1;

QY 140 TTCAGGAACCCAGGTCGGCTCTCGCGGAAATACTTCTGTTGGCTTCAACATGTT 199

DB 79 TACTCGAACCCAGGTCAGCTGCTGTACAACTACTCTCTTACAGTACAATATCGTC 138

QY 200 TTCTGGGTGCTGGAGCCCTGTTCTTGGCCATCGGCTCTGGGCTGGGGTGAAGGGT 259

DB 139 TTTTGGCTGGCTGGAGTTGTCTTCTTGGAGTCGGCTGTGGCATGGAGCGAAAGGGT 198

QY 260 GTTCTCTCCAACTCTCTGGGCTGACGATCTCGGAGGCTCGACCTGTGTGGCTGTTT 319

DB 199 GTGCTGTCCGACCTCAACAGGTGACCCGGTTGCTGGAATCGACCCCGCTGTGCTGCT 258

QY 320 GTAGTGGTGGAGGCGTCATGCTCGGCTTGGCGGCTTGGCGGCTGCATCGGGGCTCTCCGG 379

DB 259 TTGATGGTGGGCTGGTGTGATGTTACACTGGGATTCGAGGCTGTGTGGGGCCCTCGA 318

QY 380 GAGAACACTTCTCTCAAGTTTCTCAGTGTTCCTTGGGCTCACTTCTTCTCTGGAG 439

DB 319 GAGAACATCTGCTGCTCAAGTTTCTGTTGGGCGCAATGCTGCTCATCTTCTTCTCTGAA 378

QY 440 CTGGCAACAGGATCTTGGGCTTCGTATTCAAGGACTGGATTTCAGACCAGCTCAATTTC 499

DB 379 CTGGCGCTGGCGCTGTGTTGGCTTTTATTCCAAGACTGGGTGAGAGACCGGTTCCGGGAA 438

QY 500 TTCAATTAACAACAACCTCAAGGCTATCGGATGACATGACCTCCAGAACCTCATTTGAC 559

DB 439 TTCTTCGAGGCAACATCAAGTCTTATCGGATGACATCGACCTGCAGAACCTCATTTGAC 498

QY 560 TTGTCTCAGGAATATTGGTCTTCTGTCGGAGCCCGAGGGCCTAATGACTGGAACTCAAT 619

DB 499 TCCCTTCAGAAAGCTAATCAGTGTCTGGGGCTTACGGGCTTGAAGACTGGGACCTCAAT 558

QY 620 ATCTATTTCACTGCACTGACTGTAACCCGAGCCGAGAGCGCTCGGGGTGCCCTTCTCC 679

DB 559 GTCTACTTCACTGCACTGCTGCCAGCTACAGCCGAGAGAAATGTGGGGTACCCCTTCTCC 618

QY 680 TGTGTGTGAGGACCCCTGC---GATGTCTCAACACCCAGTGTGGCTATGATGTCGG 735

DB 619 TGTGTGTGCCAGATCTCTGCACAAAAGTGTGAAACACACAGTGTGGCTATGATGTCGG 678

QY 736 CTCAAACTGGAGCTGGAGCAGAGGGCTCCATACACACCAAGGCTGTGTGGGCCAGTTT 795

DB 679 ATTACGCTGAAGAGCAAGTGGGATGAGTTCACTTTTACAAAAGGATGCATCCAGGCTCTG 738

QY 796 GAGNAGTGGCTCGAGACAACTGATCGTGTGGCTGGGGCTTTTGTGGGCATCGCTCTC 855

DB 739 GAAGGCTGGCTCGCCAGGAACATCTACATTGTGGCTGGTGTCTTCAATTCCTCATCTC 798

QY 856 CTCACAGATCTTTGGTATCTGCTGGCCAGAACCTTTGTGAGTGACATCAAGGAGTGAAG 915

DB 799 CTCGAGATTTTGGCATCTTCTCTGGGAGGACCTGATCTCAGACATCAGGAGTGAAG 858

QY 916 GC 917

DB 859 GC 860

## RESULT 13

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REMARK

## COMMENT

BC024611 2498 bp mRNA linear ROD 16-APR-2003  
Mus musculus DNA segment, Chr 14, ERA70 Doi 226, expressed, mRNA  
(CDNA clone MGC:28503 IMAGE:418261), complete cds.

BC024611  
BC024611.1 GI:19354053  
MGC.

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2498)  
Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Strausberg, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullay, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krywinski, M.I., Skalska, U., Smalish, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
2238257

12477932  
2 (bases 1 to 2498)  
Strausberg, R.

Direct Submission  
Submitted (01-MAR-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleseed, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 37 Row: e Column: 12

This clone was selected for full length sequencing because it  
passed the following selection criteria: GenomScan gene  
prediction, Similarity but not identity to protein.

Location/Qualifiers  
1..2498  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strains="FVB/N"

/db\_xref="taxon:10090"  
/clone="MGC:28503 IMAGE:418261"  
/tissue\_type="Salivary gland, 10 week old female mouse"

FEATURES  
source

gene  
CDS  
BASE COUNT 542 a 644 c 631 g 681 t  
ORIGIN  
Query Match 14.5%; Score 368.8; DB 10; Length 2498;  
Best Local Similarity 67.9%; Pred. No. 3.7e-72;  
Matches 531; Conservative 0; Mismatches 247; Indels 4; Gaps 1;  
QY 140 TTCAGGACCCAGGTCGGCTGCTGGGGAATACTTCTCTTTGGCTTCAACATGTT 199  
Db 70 TACTCGAACGCCAGGTGAGTCTGCTTTCACAGTACCTCTTTCACATATATCGTC 129  
QY 200 TTCTGGGTCTGGGAGCCCTGTCTGGCCATCGGCTCTGGGCTGGGCTGAGAAGGT 259  
Db 130 TTTTGGTGGTGGAGTGTCTCTTGGAGTGGGCTGGGCTGGGATGAGCGGAAAGGT 189  
QY 260 GTTCTCTCCAAACATCTCTGCTGCTGACCGATCTGGAGGCTCGACCTCTGTGGCTGTTT 319  
Db 190 GTGCTGTCGACCTCACCAAGGTGACCCGGTTTCATGGAATCGACCCCTGCTGCTGTC 249  
QY 320 GTAGTGGTGGAGGCTCATCTGCTGGCTTGGGCTTGGGCTGATCGGGGCTCTCCGG 379  
Db 250 TTGATGGTGGGCTGCTGATGTTACACTGGGATTCGAGGCTGTGTGCGGGCCCTCCGA 309  
QY 380 GAGAACATCTTCTCTCTCAAGTCTTCTCAGTCTTCTTGGCTCTCATCTTCTCTCGGAG 439  
Db 310 GAGAACATCTGCTCTCAAGTCTTCTGTTGGGCTTCTGCTCATCTTCTTCTCTGGAA 369  
QY 440 CTGGCAACAGGATCTTGGCTTCTGATTTCAAGGACTGATTCGAGACCGAGCTCAATTTTC 499  
Db 370 CTGGCGTGGCGCTGCTGGCTTTTATTTCAAGACTGGGTGAGAGACCGGTTCCGGAA 429  
QY 500 TTCAATTAACAACAGCTCAAGCTTATCGGATGATGATGATGATGATGATGATGATGATGAT 559  
Db 430 TTCTTCGAGAGCAACATCAAGTCTTATCGGATGATGATGATGATGATGATGATGATGATGAT 489  
QY 560 TTGCTCTAGGAATATTTGCTTCTGCTGGGAGCCCGAGGCTTAACTGAGGAGGAGGAGGAGG 619  
Db 490 TCCCTTCAGAAAGCTAATCAGTCTGCGGGCTTACGGGCTTGAAGAGCTGGGAGGAGGAGG 549  
QY 620 ATCTATTTCAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 679  
Db 550 GTCTACTTCACTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 609  
QY 680 TGCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 735  
Db 610 TGCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 669  
QY 736 CTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 795  
Db 670 ATTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 729  
QY 796 GAGAG 855  
Db 730 GAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 789

Qy 856 CTCAGATCTTGGTATCTGCTGGCCAGAACCTTGTGAGTCACATCAAGGCGAGTCAAG 915  
 |||||  
 Db 790 CTCAGATTTTGGCATCTTCTGCGGAGGACCTTGATCTCAGACATCGAGGCGAGTGAAG 849  
 |||||  
 Qy 916 GC 917  
 ||  
 Db 850 GC 851

RESULT 14  
 BC026574 2500 bp mRNA linear ROD 16-APR-2003  
 LOCUS Mus musculus DNA segment, Chr 14, ERATO Doi 226, expressed, mRNA  
 DEFINITION (cDNA clone MGC:36554 IMAGE:4954874), complete cds.  
 ACCESSION BC026574  
 VERSION BC026574.1 GI:20072352  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2500)  
 Srausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Schenck, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932  
 2 (bases 1 to 2500)  
 Srausberg, R.  
 Direct Submission  
 Submitted (02-APR-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 59 Row: a Column: 4  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Genomescan gene  
 prediction.  
 Location/Qualifiers

JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REMARK  
 COMMENT

FEATURES

source 1..2500  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="MGC:36554 IMAGE:4954874"  
 /tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating  
 ductal carcinoma, 5 month old virgin mouse."  
 /clone\_lib="NCI CGAP\_Mam6"  
 /lab\_host="DH10B"  
 /notes="Vector: pCMV-SPORT6"  
 1..2500  
 /genes="D14Etd226e"  
 /notes="synonym: MGC36554"  
 /db\_xref="LocusID:52588"  
 /db\_xref="MGI:1196325"  
 50..862  
 /codon\_start=1  
 /product="tetraspanin similar to TM4SF9"  
 /protein\_id="AAH26574.1"  
 /db\_xref="GI:20072353"  
 /db\_xref="LocusID:52588"  
 /translation="MHYRYSNAEVSCHWYKLLFSNIVFWLAGVFLGVLWAMSEK  
 GVLSDLTQVTRLHGDVPLVLMGVMTFLFAGCVGALRENICLLKFCGAILFIF  
 FLELAVLAFQDWRDRFRFESNFKSYRDDIDLNLIDSLQKANCOCGAYGE  
 DMDLVNVCSGASYSRKGVPFSCVDPDPAKVNTQCGYDVRQLKSKWBEFIF  
 KSCIQALEGLPRNIVIVAGVPIAISLLQIFGLARTLISDIEAVKAGHF"  
 BASE COUNT 545 a 643 c 631 g 681 t  
 ORIGIN  
 Query Match 14.5%; Score 368.8; DB 10; Length 2500;  
 Best Local Similarity 67.9%; Pred. No. 3.7e-72;  
 Matches 531; Conservative 0; Mismatches 247; Indels 4; Gaps 1;  
 Qy 140 TTCAGAAACCCGAGTCCGCTGCTCGGGAAATATCTCTGTTGGCTTCAACATTGTT 199  
 Db 65 TACTCGAACCCGAGGTGAGTCTGCTGATCAAGTACTGCTCTTACGTAACAATATCGTC 124  
 Qy 200 TTCTGGGTGCTGGAGCCCTGTTCTCGGCATCGGCTCTGGGCTGGGGTGGAGAGGTT 259  
 Db 125 TTTTGGCTGGTGGAGTTGTTCTTCTTGGAGTCGGGCTGTGGGCATGGAGCGAAAGGGT 184  
 Qy 260 GTTCTCTCAACATCTCTCGCTGACCGATCTCGGAGGCTCGACCTGTGTGGCTGTTT 319  
 Db 185 GTGCTGCGACCTCAACAGGTGACCCGGTTGCATGGATCGACCCGCTGCTGCTGTC 244  
 Qy 320 GTAGTGGTGGAGGCTCATGTCCGTGCTGGGCTTTGCCGGTGCATCGGGGCTCTCCGG 379  
 Db 245 TTGATGGTGGGCTGGTGTGATGTTACACTGGGATTCGACGGCTGTGTGGGGCCCTCCGA 304  
 Qy 380 GAGAACACTTCTGCTGCTCAAGTTTCTCAGTGTCTTGGGCTCATCTTCTTCTGGAG 439  
 Db 305 GAGAACATCTGCTGCTCAAGTTTCTTGTGGGGCCATTGTGCTCATCTTCTTCTCGGAA 364  
 Qy 440 CTGGCAACAGGATCTTGGCTTTCGTTATTCAGAGACTGGATTTCGAGACCAAGCTCAATTC 499  
 Db 365 CTGGCGCTGGCGCTGTGTGGCTTTTATTCAGAGACTGGTGGAGAACCGGTTCCGGAA 424  
 Qy 500 TTCAATTAACAACACTCAAGGCTATCGGGATGACATTTGACCTCCAGAACCTTCTGAC 559  
 Db 425 TTCTTCGAGAGCAACATCAAGTCTTATCGGATGACATCGACCTGCAGAACCTTCTGAC 484  
 Qy 560 TTGTCTCAGGAATATGGTCTTGTCTGGAGCCGAGGGCTTAATGACTGGAACTCAAT 619  
 Db 485 TCCTTTCAAGAGCTAATCAGTGTCTGGGGCTTACGGCCCTGGAAGACTGGGACCTCAAT 544  
 Qy 620 ATCTATTTCACCTGCACTGACTTTGAACCCGAGCGGAGCGCTGGGGTGGCTTCTCC 679  
 Db 545 GTCTACTTCACTGCAAGTGGTGGCAGCTACAGCCGAGAGAAATGTGGGTACCTTCTCC 604  
 Qy 680 TGTGTGTGAGGAGCCCTGC----GATGTCTCAACACCCAGTGTGGCTATGATGTCCGG 735  
 Db 605 TGCTGTGTGCCAGATCTCTGCACAAAAAGTCGTGNAACACACAGTGTGGCTATGATGTCCGG 664

Qy	736	CTCAAACTGGAGCTGGACGACGAGGCTCCATACACAAAGGCTGTGTGGGCCAGTTT	795
Db	665	ATTCAGCTGAAGACGACGTGGGATGAGTTCATCTTTACAAAGGATGCATCCAGGCTCTG	724
Qy	796	GAGAACTGGCTGCAGGACAACTGATCTGTGTGGCTGGGGTCTTTGTGGGCATCGCTCTC	855
Db	725	GAAGCTGGCTGCCAGGAACATCTACATTTGGCTGGTCTTCATTGCCATCTCACTG	784
Qy	856	CTCAGATCTTTGGTATCTGCCTGGCCAGAACCTTTGTAGTGACATCAAGGCGATGAAG	915
Db	785	CTGCAGATTTTGGCATCTTCTGGCGAGGACCCCTGATCTCAGACATCGAGGCGATGAAG	844
Qy	916	GC 917	
Db	845	GC 846	

RESULT 15  
AF3111903  
LOCUS  
DEFINITION Homo sapiens DC-TM4F2 precursor, mRNA, complete cds.  
ACCESSION AF3111903  
VERSION AF3111903.1 GI:22266721  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Zhang, W., Li, N., Wan, T. and Cao, X.  
TITLE Identification of novel membrane proteins  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2553)  
Zhang, W., Li, N., Wan, T. and Cao, X.  
AUTHORS Direct Submission  
TITLE Submitted (06-OCT-2000) Department of Immunology, Second Military  
Medical University & Shanghai Brilliant Biotechnology Institute,  
800 Xiangyin Rd., Shanghai 200433, P.R.China

```

location=VGAarrilels
1. .2553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
73. .885
/codon_start=1
/product="DC-TM4F2 precursor"
/protein_id="AAM94899.1"
/db_xref="GI:22266722"
/translations="MHYRYSNAKVCWYKYLLFSYNIFWLGVVFLGVLMAWSEK
GVLSDLTKYRMHGIDPVVLVLVMGVVMTLGPAGCVGARNICLLNPFPGTIVILP
FLELAVALIAFLQDWRDRFRFFESNIKSRYRDDIDLQNRLDSLRANQCAGAYGE
DWLNLYFNCSGASYGREKGVFPSCVPDPAPKVNTQCGYDVRTQLSKKWDESIPT
KGCIQAELSPRNIVYAVGVFTAISLLQLFIGIFLTARLTISDIEAVKAGHFH"
505 a      724 C    680 G    644 t
BASE COUNT
ORIGIN

```

	Query Match	13.9%	Score 353.4	DB 9	Length 2553	
	Best Local Similarity	67.0%	Pred. No. 1.1e-68			
	Matches 518	Conservative 0	Mismatches 251	Indels 4	Gaps 1	
QY	150	CCGAGTTCGGCTCTCGCGGAATAC	TCTCTGTTGGCTTCAACATTGTTTTCGGGTGC	209		
Db	98	CCAAGGTTCAGCTGCTGTGTACAAGTACCTCTCTTTCAGCTACACATCATCTTCTGTTGG	157			
QY	210	TGGGAGCCCTGTTCTCTGGCCATCGGCCTCTGGGCTCGAGAGGGTGTTCTCTCCCA	269			
Db	158	CTGGAGTTGTCCTCTTGGAGTCGGGCTCTGGGCATGAGCGAAAGGTGTGCTGTCCG	217			
QY	270	ACATCTCTGGCGCTGACCGCATCTGGGAGGCCTCGACCCCTGTGTGGCTGTTTGTAGTCGTTG	329			
Db	218	ACCTCAACAAAGTGACCCCGGATCGATCGAATCGACCCCTGTGGTCTGGTCTCATGTGTGG	277			

Qy	330	GAGCGGTCAATGTCGTGCTGGGCTTTGGCCGGCTGCATCGGGGCTCTCGGGGAGAACACTT	388
Db	278	CGTGGTGATGTTACCCCTGGGGTTGCGCCGCTGCGTGGGGGCTCTCGGGGAGAAATATCT	337
Qy	390	TCCTGCTCAAGTGTTTTCTCAGTGTTCCTTGGCCCTCATCTTCTTCCTGGAGCTGGCAACAG	449
Db	338	GCTTGCTCAACTTTTTCGTGGCACCATCGTGTCTCATCTTCTTCCTGGAGCTGGCTGTGG	397
Qy	450	GGATCTTTGGCCTTCGTATTCGAAGGACTGGATTCGAGACCAGCTCAATTTCTTCATTAACA	509
Db	398	CCGTGCTGGCCTTCCTGTTCCAGGACTGGGTGAGGACCGGTTCCGGGAGTTCTTCGAGA	457
Qy	510	ACAACGTCAAGGCCATTCGGGATGACATTGACCTCCAGAACTCATTTGACTTTGCTCAGG	569
Db	458	GCAACATCAAGTCTACCGGAGCATATCGATCTGCAAAACCTCATCGACTCCCTTCAGA	517
Qy	570	AATATTGGTCTTGTGTCGGAGCCCGAGGGCCTAATGACTGGAACTCAATATCTATTTCAC	629
Db	518	AAGCTAACAGTGTGTGTGGCGCATATGGCCCTGAAGACTGGGACCTCAAGCTCTACTTCA	577
Qy	630	ACTGCACTGACTTGAACCCGAGCCGAGAGCGCTCGGGGTGCCCCCTTCCTGCTGTGTCA	689
Db	578	ATTGACGGGTGCCAGCTACAGCCGAGAGAAGTGGGGGTCCCCCTTCTCCTGCTGGTGC	637
Qy	690	GGGACCTGGG----ATGTCTCTCAACACCCAGTGTGGCTATGATGTCCGGCTCAAACTGG	745
Db	638	CAGATCTCGGCAAAAAGTTGTGAACACACAGTGTGGATATCATGTCAAGGATTCAGCTCA	697
Qy	746	AGCTGGAGCAGCAGGGCTCCATACACACCAAGGCTGTGGGCCAGTTTCGAGAAGTGGC	805
Db	698	AGAGCAAGTGGATAGTGTCCATCTTCAGAAAGGCTGCATCCAGGCGCTGGAAAGCTGGC	757
Qy	806	TGCAGGACAACTGATCGTGTGGCTGGGGTCTTTTGTGGGCATCGCTCTCCTCCAGATCT	865
Db	758	TCCCGCGGAACATTTACATTGTGCTGGCGTCTTCATCGCCATCTCGCTGTTTGCAGATAT	817
Qy	866	TTGGTATCTGCTGGGCCAGAACTTGTGAGTGAACATCAAGGCAAGTGAAGGCC	918
Db	818	TTGGCATCTTCTCGCAAGGACGCTGATCTCAGACATCGAGGCAAGTGAAGGCC	870

Search completed: November 21, 2003, 21:30:42  
Job time : 9172 secs